THIS WEEK

EDITORIALS

ASTRONOMY The comfort of strange supermassive black holes **p.152**

WORLD VIEW Scientists must embrace politics and politicians **p.153**



BIODIVERSITY Amazon in Ecuador jumping with frog species **p.154**

The press under pressure

With the Leveson inquiry scrutinizing journalistic practice in the United Kingdom, scientists should take the opportunity to fight back against agenda-driven reporting.

Everyone has an example of the scientific ignorance of the press, but researchers in Britain probably have more than most. With stories ranging from ludicrous (wind turbine attacked by aliens) to downright irresponsible (promoting the link between childhood vaccinations and autism), the fourth estate in the United Kingdom has hardly covered itself in glory when it comes to science and scientific issues.

Other countries have similar grievances, of course — particularly the United States, where right-wing talk radio and cable television regularly air anti-science views on everything from global warming to creationism. Stem-cell scientists in Germany and transgenic-crop researchers in France have also been assailed by journalism out of step with the scientific evidence that it claims to examine. But there is a sense that the situation is more acute in tabloid-driven Britain, particularly given the distasteful news-gathering techniques that are now under the microscope like never before.

In Britain, eyes are on an inquiry into the standards and ethics of the press, headed by Lord Justice Brian Leveson. Widely known as the 'phone-hacking inquiry', triggered as it was by revelations about the extent of illegal eavesdropping at the now-defunct *News of the World* newspaper, the judicial investigation in fact has a much wider scope. In his opening remarks, Robert Jay QC, counsel to the inquiry, said that he expected members of the scientific community to submit evidence that sections of the press were causing real harm by not basing their commentaries on evidence and not applying the scientific method to their reports. But his remarks seem to have surprised many within the community he was referring to — a subsequent search by the Science Media Centre, an advocacy group in London, found just a single planned submission, and it now intends to send its own.

This is a wasted opportunity. Too often, talk about the difficult relationship between the media and science gets bogged down in well-meaning but ultimately naive discussion of how to 'help' reporters to get their facts straight. Should journalists send stories to scientists to be vetted before publication? Should they have scientific training? Should scientists be trained to offer sound bites?

Alastair Campbell, the former communications chief for the UK government and a man who knows a thing or two about sound bites, better characterized the problem when he told the Leveson inquiry about "agenda-driven journalism regardless of facts". Campbell was referring to the media fixation with autism and childhood vaccinations, but it could have been anything from climate change (on which positions range from 'it's not dangerous' to 'it's going to kill us all') or volcanic ash and aviation (it's not dangerous) to bird flu (it's going to kill us all).

Sometimes the agenda is obvious and explicitly political. More often, it is the instinctive overreach of a story-teller who chooses what to include to make their tale as interesting as possible. Either way, the problem runs deeper than reporters regularly confusing bacteria and viruses, however irritating that may be to some. Witness the list of the top 100 UK political journalists of 2011, as decided by *Total Politics* magazine last week. In at number 14 is Christopher Booker, a columnist for *The Sunday Telegraph*. Not satisfied with Booker being "skeptical about global warming", the magazine partly credits his high rank to his claims that evidence to prove that passive smoking and exposure to asbestos cause cancer "does not exist".

Journalism that favours attitude over accuracy is more common than scientists suspect, and not just on the comment pages or in the tabloids. And it is also more damaging — with news editors behind the scenes ordering certain lines on high-profile stories, no matter what the science says.

"Journalism that favours attitude over accuracy is more common than scientists suspect."

The evidence offered to Lord Leveson shows that science is far from alone in this treatment: parents of missing and murdered children have queued up to tell harrowing stories of blatant misrepresentation, on top of the indignity of having voicemails hacked.

But science has a way to respond that others do not. Through online forums, blogs and Twitter, a cottage industry has grown up around instant criticism of dodgy scientific claims and dubious findings. This parallel journalism is increasingly coming to the attention of the mainstream press — as demonstrated by the rising number of stories in the press that were first broken by blogs.

It may seem thankless at times, but the army of online commentators who point out the errors, the inconsistencies and the confounding factors, and from time to time just scream 'bullshit', have the power to hold the press to account. This ongoing war of attrition against those who would put their own agendas above the facts cannot take away their platform, but it can chip away at something they prize even more: their relevance, and with it their pernicious influence. ■

Troubled waters

Protected areas are only the start of the road to reforming our relationship with the seas.

The easiest way to create a nature reserve from a car park is simply to declare it as such. The land is then designated as protected, and counts towards the relevant government's targets to set aside a certain amount of its territory from development. That is a ridiculous example, of course, and would never happen on land — so why do we allow a similar exercise to happen in the sea?

No one should doubt that our seas need protection. Overfishing, pollution and climate change are fundamentally changing some of the

© 2011 Macmillan Publishers Limited. All rights reserved

most important regions on the planet. And the conservation response — marine protected areas (MPAs) — should be a key tool to safeguard the world's maritime environment. By setting aside areas in which human activity is tightly regulated, the thinking goes, governments can ensure that key habitats and species are preserved.

Certainly, good MPAs such as those in Hawaii and around the Great Barrier Reef can bring great benefits, from saving individual species from extinction to preserving entire coral-reef or deep-sea ecosystems. And the number of these areas is on the rise, as the world's governments crawl towards a target to protect 10% of the oceans.

But to get the most from these areas, much greater efforts are needed. In the sea, as on land, balancing the recommendations of scientists with commercial needs and other interests is always tricky for politicians. Still, as highlighted on page 166, and as shown by recent developments in Australia (see *Nature* **480,** 14–15; 2011), there are growing concerns about the poor science base of many of the existing and proposed MPAs.

As Australia divides up its waters into a patchwork of protected, recreational, commercial and multipurpose areas, researchers say that the country is ignoring evidence that would make this zoning stronger. These concerns, which surfaced over proposals for the southwest region, again in the northwest region and then most recently in the plan for a huge reserve in the Coral Sea, centre on whether enough sea has been given the highest level of protection and whether the areas that are highly protected include a representative selection of all necessary habitats. The second question in particular is one on which the voices of researchers must be heard by government, and acted on.

And even when reserves are established on the basis of sound science, governments must ensure that they can protect the areas properly. It is telling that even the prosperous United States may be failing in this. Jane Lubchenco, the head of the US National Oceanic and Atmospheric Administration, says: "We don't have the resources that we need to actually monitor, enforce and understand these areas." If the United States is having trouble, what hope is there for the less-developed nations?

"Declarations of protection are worthless without actions and, crucially, without money."

In the sea, as on land, declarations of protection are worthless without actions and, crucially, without money.

Equally, governments must not fall into the trap of thinking that designating and protecting a certain percentage of our seas will solve all maritime conservation problems. Larger efforts to reform our approach to the oceans, from fisheries to extraction, are long overdue. Efforts in New England waters have demonstrated that it might be possible to manage local fisheries sustainably (see *Nature* **465,** 540–542; 2010), and international agreements — such as those that have safeguarded Antarctica — show that countries can pull together under the right circumstances. For too long what has been out of sight has also been out of mind. The rise of MPAs should be a starting point for a reformation, not a tool to hit a target so that we say 'job done' and move on to other issues. ■

A massive comfort

There are bigger things in the Universe than our earthly worries. Much bigger.

People have always looked to the heavens for inspiration and distraction from everyday life. Some see their futures in the stars and others talk to the man in the Moon. But for the really big problems, you need a really big distraction. So for readers who feel especially gloomy, perhaps about the latest bout of political shadowboxing on global warming in Durban, South Africa, or the perilous state of European finances, let *Nature* introduce you to two newly identified black holes. These are not the ordinary, two-a-penny stellar black holes, mind, but their supermassive relatives — the big beasts of the cosmos.

On page 215 of this issue, astronomers from across the United States present their discovery of the two most massive black holes ever found. The larger of the pair would tip the scales at some ten billion times the mass of our own Sun. Heavy stuff indeed.

In some respects, it is surprising that these black holes have managed to hide for so long. The curious twinkling of some quasars dated to the earliest days of the Universe had made astronomers pretty sure that the giants were out there somewhere, yet the previous heavyweight champion, in the giant elliptical galaxy Messier 87, is a relatively puny 6.3 billion solar masses.

Measuring the motion of stars from two different galaxies that are the biggest in their clusters, NGC 3842 and NGC 4889, Nicholas McConnell of the University of California, Berkeley, and his colleagues found that NGC 3842 has a central black hole with a mass of 9.7 billion solar masses, and NGC 4889 has a black hole about the same size or even larger. (All massive galaxies with a spheroidal component, such as the bulge of the Milky Way, are thought to harbour a supermassive black hole at their centre.)

As celestial inspirations go, it is true that the giant twins of darkness at the heart of NGC 3842 and NGC 4889 are somewhat distant companions. The nearest is about 98 megaparsecs from Earth — about 320 million light years. And it is true that McConnell's team was able to look to the sky with more than a shrug of resignation — the researchers could use integral-field spectrographs at the giant Gemini North and Keck telescopes perched on Hawaiian mountains. But what these supermassive black holes lack in touchy-feely access, they surely gain in being really, really big. And if it helps those of us weighed down by terrestrial matters, they are out there, somewhere.

As supermassive black holes go, they are perhaps just a little too supermassive, certainly a bit larger than astronomers would have predicted given what they know about their host galaxies. Perhaps the processes that influence the growth of the largest galaxies and their black holes are somehow different from those in smaller galaxies? As the authors note, better adaptive optics instruments on telescopes and very-long-baseline radio interferometry are finding black holes in ever more exotic galaxies, and will help to answer that question.

It now seems plausible that, within a decade or so, astronomers will be able to capture the first direct image of a black hole. Plans are afoot to link a network of instruments, stretching from the high desert plains of Chile and the South Pole to potential sites in Africa and New Zealand. Pointed at the centre of our own galaxy, such a telescope should be able to see SgrA*, the supermassive black hole believed to sit there. Astronomers already have an image in mind: a bright ring spun around a dark shadow cast by the black hole, because no light escapes its event horizon.

Widely described as the point of no return, the event horizon of SgrA* would be the largest in our skies, but still just 30 microarcseconds across — the apparent size of a tennis ball on the Moon when viewed from Earth. To capture its image would be a stunning technical achievement in itself, but it would also open the door to further studies of how black holes spin and gather material, as well as probing some fundamental aspects of space-time and general relativity. And the first picture taken of our local supermassive black hole — the most enigmatic and charismatic of all the wonders of the Universe — would surely be one of the defining images of the time. It might even knock everyday trouble and strife from the front pages, and perhaps even, for a while, from people's minds. ■

➲ NATURE.COM
To comment online, click on Editorials at:
go.nature.com/xhunqv

© 2011 Macmillan Publishers Limited. All rights reserved

SEVEN DAYS *The news in brief*

POLICY

Europe budget rise

The European Commission has proposed an €80-billion (US$107.5-billion) budget for Europe's next research funding programme, which runs from 2014 to 2020. That is a 60% rise on the current programme (running 2007–13), but individual sectors get larger funding boosts: the European Research Council, a grants agency for basic research, is in line for a 77% rise to €13.2 billion. Europe's research commissioner, Máire Geoghegan-Quinn, emphasized that stem-cell research would be supported, assuaging fears that it might not after Europe's Court of Justice banned patents on human embryonic stem cells in October. The proposals, released on 30 November, will now be discussed by the European parliament and by member states, with the aim of reaching an agreement by the end of 2013. See go.nature.com/hxmqua for more.

UK medical boost

British scientists may gain wider access to data from patients in the country's National Health Service, the government announced on 5 December. It is consulting on plans for patient data to be automatically 'opted in' to clinical research, which could include research by private companies. The consultation is part of a package of measures to boost Britain's life-sciences industry, including a fund (half of it new spending) worth £180 million (US$282 million) to speed drugs from lab to clinic. See go.nature.com/qe6wmp for more.

ITER budget saved

After more than a year of deadlock, European policy-makers have agreed a deal to plug a €1.2-billion (US$1.6-billion) funding gap in the 2012–13 budget of ITER, the €15-billion experimental fusion reactor being built near Cadarache, France. The agreement redirects €840 million from other areas of European spending, and importantly avoids raiding Europe's research budget, which ministers had threatened to do. The remaining €360 million will come from as-yet-unspecified areas of Europe's 2013 budget, policy-makers agreed. See go.nature.com/qjnxy3 for more.



S. VAN ZUYDAM/AP

Durban debates climate treaty

Delegates at the United Nations climate talks in Durban, South Africa, are asking afresh whether the world needs a new legally binding climate treaty when the Kyoto Protocol expires next year. Campaigners rallying in the city (pictured) urged a breakthrough, but none had arrived during the first week of talks.

Negotiators did at least make some progress on initiatives to curb deforestation, striking a technical deal that lays out ground rules by which countries would report baselines for rates of deforestation, and would submit these to international peer review. See go.nature.com/gmzoab for the latest blogs from Durban.

Plutonium to burn

Hoping to get rid of its 112-tonne stockpile of plutonium, the UK government said last week that it was considering building a plant to recycle the plutonium into mixed oxide fuel (MOX). The move comes just four months after the Nuclear Decommissioning Authority announced the closure of the country's only nuclear reprocessing facility, at Sellafield (see *Nature* **476,** 140; 2011). That plant (one of only two commercial MOX facilities in the world) has been beset with technical problems. But a new MOX reprocessing plant is not the only option: on 30 November, GE Hitachi Nuclear Energy, based in Wilmington, North Carolina, proposed building a fast reactor fuelled by plutonium. See go.nature.com/lrsqza for more.

RESEARCH

Elements named

Pencil two new names into the periodic table: elements 114 and 116 should be called flerovium (Fl) and livermorium (Lv), the International Union of Pure and Applied Chemistry (IUPAC) proposed on 1 December. The names honour the institutes that collaborated to create the artificial elements: livermorium for the Lawrence Livermore National Laboratory in California, and flerovium for Georgy Flerov, founder of the Flerov Laboratory of Nuclear Reactions at the Joint Institute for Nuclear Research in

© 2011 Macmillan Publishers Limited. All rights reserved

Dubna, Russia. (Dubnium is already an element.) IUPAC will allow five months for public comment, but the names are likely to be confirmed next May. See go.nature.com/2wvwkc for more.

Earth-like planet

Kepler, NASA's planet-hunting space telescope, has announced its first potentially habitable planet around a star like our Sun. Kepler-22b is not an Earth twin: at 2.4 times Earth's radius, it could be more like a gassy Neptune than a rocky planet. But it sits in a comfortably warm orbit from its host star, at temperatures similar to those experienced by Earth. The planet — announced on 5 December — is 184 parsecs (600 light years) from us. Kepler's full catalogue now includes 2,326 planet candidates. See go.nature.com/xc6we6 for more.

Stapel retraction

Dutch social psychologist Diederik Stapel, who admitted faking data in numerous published papers, last week retracted a 2011 study that suggested disorderly environments encourage discrimination (D. A. Stapel and S. Lindberg *Science* **332,** 251–253; 2011). It is the first in what is likely to be dozens of retractions. Committees at the universities of Amsterdam, Groningen and Tilburg, where Stapel studied and worked between 1994 and 2011, aim to identify tainted papers for a report that will not be finished until mid-2012. See go.nature.com/8lv4n2 for more.

Warmest La Niña

This year is set to be the tenth warmest since records began in 1850, the World Meteorological Organization (WMO) said on 29 November. But it will also be the hottest year to have experienced a strong La Niña event — the natural cycle in which temperature falls over vast areas of the sea surface in the equatorial east-central Pacific Ocean. The WMO used figures up to October from three data sources to make its projection.

Open science apps

A website where people can upload their genetic and phenotypic information has come first in a contest to build applications to make science more open. OpenSNP (opensnp.org) won the inaugural API Binary Battle, sponsored by open-access publisher the Public Library of Science and the paper-sharing site Mendeley. See go.nature.com/wdw9gw for more.

Deforestation down

Deforestation in the Brazilian Amazon rainforest has fallen to a record low for the third year



J. STANMEYER/VII/CORBIS

running, according to satellite data released on 6 December by Brazil's National Institute for Space Research (INPE) in São José dos Campos. It seems that strong government enforcement countered forest clearing after sudden spikes in deforestation (**pictured**) were reported from March to May. In August, the INPE had reported a 15% rise in deforestation for the year, which runs from August 2010 to July 2011. But that report was based on coarser-resolution satellite imagery. See go.nature.com/pxpkyg for more.

BUSINESS

Arctic oil setback

Cairn Energy, a company that is leading efforts to find oil and gas off the coast of Greenland, announced on 30 November that it is plugging and abandoning the final wells it drilled this year. The firm, based in Edinburgh, UK, has failed to find any commercially viable reserves after two years drilling eight offshore exploration wells at a reported cost of about US$1 billion. Market analysts speculated that the company might give up drilling in 2012, although other companies, including Shell, Exxon, Chevron and Statoil, hold licences to explore off Greenland.

Meltdown admitted

The operators of Japan's stricken Fukushima Daiichi nuclear plant have made their direst assessment yet of the meltdown that took place in March. In the most damaged reactor, Unit 1, all the fuel melted through the steel pressure vessel and seeped into (but not through) the concrete base of the containment vessel below, the Tokyo Electric Power Company (TEPCO) admitted on 30 November. TEPCO had previously said that only some of the fuel had melted, but its latest simulations fall into line with models run by observers such as the US Nuclear Regulatory Commission. Leaks of radioactive water continue at the site: on 4 December, TEPCO said that it was using sandbags to contain a new leak from a building in which water was being decontaminated.

COMING UP

9 DECEMBER
International climate negotiations in Durban, South Africa, conclude. See go.nature.com/dw9pu2 for the latest from the meeting.

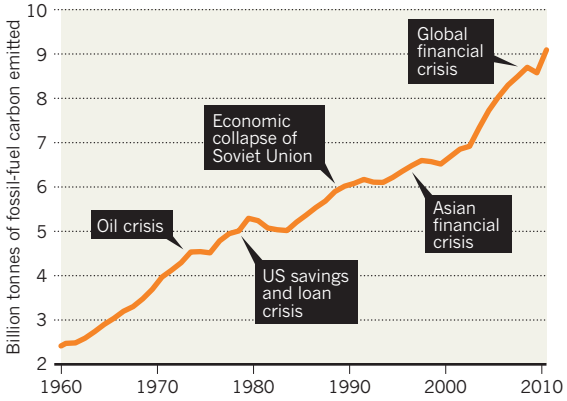
13 DECEMBER
Latest results from the Large Hadron Collider will be revealed at a CERN public seminar in Geneva, Switzerland. See go.nature.com/lyy96r for rumours before the event.
go.nature.com/rqgjuc

TREND WATCH

The sharp rise of carbon emissions from 2009 to 2010 — by nearly 6% — suggests that the global financial crisis has had little impact in slowing down the world's output of greenhouse gases, according to an analysis published on 4 December (G. P. Peters *et al. Nature Climate Change* http://doi.org/g2m; 2011). The rise was mainly due to high emissions growth in a few key economies. China's emissions were up by 10.4%, India's by 9.4% and Brazil's by 11.6%. See go.nature.com/bvlgbj for more.

CARBON SHRUGS OFF CRISIS

Global carbon emissions seem to have recovered faster from the global financial crisis than from previous economic crises.



SOURCE: NATURE CLIMATE CHANGE

NATURE.COM
For daily news updates see:
www.nature.com/news

© 2011 Macmillan Publishers Limited. All rights reserved

NEWS IN FOCUS

GEOSCIENCE Cooking up a model of Earth's core in the laboratory **p.162**

EDUCATION Will cuts sacrifice excellence at the University of California? **p.164**

OCEANOGRAPHY China's ambitions ride a spanking new ship **p.166**



PROFILE Henrik Ehrsson probes the mind with body illusions **p.168**

AP PHOTO/H. N. GHANBARI



US President Barack Obama last week pledged to fund HIV treatment for millions more people worldwide over the next two years.

GLOBAL HEALTH

Cutbacks threaten HIV gains

Wealthy nations have reduced contributions to global research and treatment programmes.

BY MEREDITH WADMAN

Thirty years after AIDS was first recognized as a human scourge, major recent gains in treatment and prevention risk being derailed by the global financial crisis.

On 23 November, the Global Fund to Fight AIDS, Tuberculosis and Malaria announced that it will not fund new grants for prevention and treatment until 2014, owing to "substantial budget challenges in some donor countries". The fund's HIV activities run the gamut from counselling and testing pregnant women in India to providing medications to infected children in Kenya.

And on 7 December, a report from Policy Cures in Sydney, Australia, a group that monitors global research and development for neglected diseases, showed that public and private funders last year cut their commitments to HIV/AIDS research targeted at the developing world by US$67.5 million, or 5.9%. The decline was due entirely to cutbacks by wealthy nations, which slashed spending by $72.6 million; poorer countries actually increased funding by $5.1 million. Where research is concerned, "AIDS had a bigger drop in dollar terms than any other disease", says Mary Moran, the executive director of Policy Cures and the lead author on the report.

The threat comes even as US President Barack Obama last week promised that the United States will step up its worldwide attack on AIDS. Breaking the White House's past reticence on the issue, he urged nations to honour their unmet pledges to the Global Fund.

"Countries that have committed to the Global Fund need to give the money that they promised," he said on 1 December, World AIDS Day. "Countries that haven't made a pledge, they need to do so," he added, singling out emerging nations such as China that are recipients of funds "but now are in a position to step up as major donors".

NATURE.COM
For more, see *Nature*'s Outlook on HIV/AIDS:
go.nature.com/ms4u9t

The United States has given $6.1 billion to the Global Fund since 2004, and last year pledged to contribute $4 billion between 2011 and 2013. Congress approved $1.05 billion in 2011, meaning that allotments must increase substantially in 2012 and 2013 to honour the pledge. Obama has asked Congress to provide $1.3 billion in 2012, but Senate lawmakers have so far resisted the increase.

Obama pledged to increase by 50% — to 6 million — the number of people receiving antiretroviral therapy (ART) worldwide by 2013 through the President's Emergency Plan for AIDS Relief (PEPFAR), the nation's major global treatment and prevention programme. He also said PEPFAR would aim to provide ART to an additional 1.5 million pregnant women with HIV in the next two years. The international promises came with no new money immediately attached. But Obama has asked Congress to provide nearly $7.2 billion — a 6% increase — for PEPFAR in 2012, part of the ▶

© 2011 Macmillan Publishers Limited. All rights reserved

unfinished budget bills still being debated.

Advocates said that his speech marked a significant turning point. "This is the first time he's signalled that he's going to champion global HIV in this way," says Jennifer Cohn, an assistant professor of infectious diseases at the University of Pennsylvania in Philadelphia, and a policy adviser in Nairobi for Médecins Sans Frontières (MSF; also known as Doctors Without Borders). "Whether or not this gets translated into the president's 2013 budget request, or what Congress chooses to protect during the 2012 budget negotiations — that's what I'm waiting to hear."

Even if the United States manages to expand treatment for HIV/AIDS, research and development aimed at poor countries is suffering. Moran points out that of the $67.5 million cut from HIV/AIDS research spending in 2010, $32 million was due to cuts by the US National Institutes of Health, and a further $17 million was cut by the UK Department for International Development (see 'Fighting a killer'). Other neglected diseases are seeing similar reductions in research funding (see go.nature.com/sgv7oj).

"A lot of this reflects the financial crisis, that's undoubted," Moran adds. "But with HIV, some of it also reflects a shift from the more costly late-stage vaccine and microbicide trials that haven't worked. We are moving back to basic research, which is cheaper."

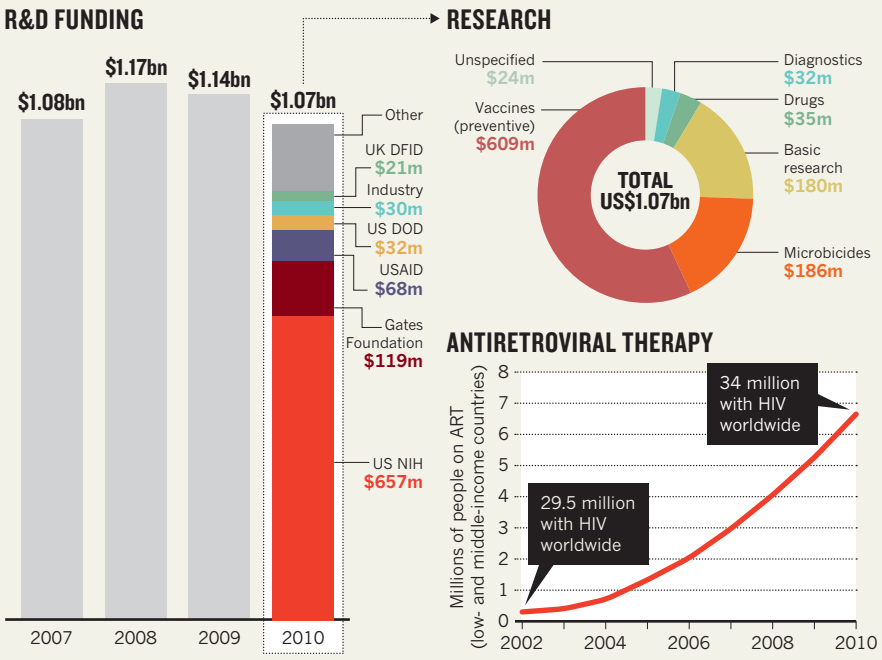
The cutbacks coincide with encouraging results on HIV/AIDS treatment. One study published in August found that early treatment of those infected greatly decreases sexual transmission to their uninfected partners, (M. S. Cohen *et al. N. Engl. J. Med.* **365,** 493–505; 2011). Other recent work suggests that ART, although costly, will pay for itself and more by increasing labour productivity, reducing the need for orphan care and deferring medical treatment costs (S. Resch *et al. PLoS ONE* **6,** e25310; 2011). "Not only is [treatment] the right thing to do, it's also the smart thing to do," says Sharonann Lynch, an HIV policy adviser with MSF in New York.

But the 'treatment equals prevention' equation invoked by Obama is not selling in all quarters. Roger Bate, a health economist and HIV expert at the American Enterprise Institute, a think tank in Washington DC, says that he is not yet persuaded by the studies suggesting that treatment prevents transmission. "There is a lot of effort out there to ignore noise in those studies," he says.

He adds that, even if treatment does prevent transmission, the commitment to treat infected people for their lifetimes quickly becomes an enormous financial outlay in future years. "Is it sustainable to treat even more people under current budgets?" he asks. "I don't think it is." ■

FIGHTING A KILLER

Funding for research and development (R&D) of HIV/AIDS products targeted at developing countries fell last year, whereas treatment with antiretroviral therapies has grown rapidly over the past decade.



SOURCE: POLICY CURES

Figures are adjusted for inflation and reported in 2007 US dollars. ART, antiretroviral therapy; DFID, Department for International Development; DOD, Department of Defense; NIH, National Institutes of Health; USAID, US Agency for International Development.

NANOTECHNOLOGY

Nano rules fall foul of data gap

Incomplete nanotoxicology research is hampering efforts to introduce regulation.

BY EUGENIE SAMUEL REICH

The science of the very small will lurk in many a Christmas gift this year. Nanomaterials add strength to golf clubs, odour resistance to socks and ultraviolet protection to cosmetics. But some of their properties could also pose health risks, and regulations covering their manufacture and use have failed to keep pace with the rush to market.

With sizes measured in mere billionths of a metre, nanomaterials have very high surface area relative to their mass, and this could alter their toxicity compared with the same material in bulk. Yet many nanomaterials are not regulated separately — in part because the high-quality nanotoxicity research needed for regulation is lacking, as an assessment presented at last week's Materials Research Society meeting in Boston, Massachusetts, made clear.

Groups representing food manufacturers had asked Cantox Health Sciences International, a consultancy in Mississauga, Canada, to review the literature for studies that could shed light on the effects of consuming nanoparticles from food packaging, in which they may be used to improve strength or even detect spoilage.

Cantox sampled 30 papers describing studies in which nanoparticles were fed to animals or applied *in vitro* to cells of the gastrointestinal tract. Of 21 animal studies, the Cantox researchers found that 15 failed to attain the highest grade on the Klimisch test, a quality-control criterion that assesses whether basic data, such as the number or gender of animals used, have been reported. "Some didn't give the dosage. This is a fairly important aspect," says Jeffrey Card, a toxicologist at Cantox who led the study.

The underlying materials science was equally spotty. Of 10 parameters that Cantox concluded should be reported in a nanotoxicity

© 2011 Macmillan Publishers Limited. All rights reserved

unfinished budget bills still being debated.

Advocates said that his speech marked a significant turning point. "This is the first time he's signalled that he's going to champion global HIV in this way," says Jennifer Cohn, an assistant professor of infectious diseases at the University of Pennsylvania in Philadelphia, and a policy adviser in Nairobi for Médecins Sans Frontières (MSF; also known as Doctors Without Borders). "Whether or not this gets translated into the president's 2013 budget request, or what Congress chooses to protect during the 2012 budget negotiations — that's what I'm waiting to hear."

Even if the United States manages to expand treatment for HIV/AIDS, research and development aimed at poor countries is suffering. Moran points out that of the $67.5 million cut from HIV/AIDS research spending in 2010, $32 million was due to cuts by the US National Institutes of Health, and a further $17 million was cut by the UK Department for International Development (see 'Fighting a killer'). Other neglected diseases are seeing similar reductions in research funding (see go.nature.com/sgv7oj).

"A lot of this reflects the financial crisis, that's undoubted," Moran adds. "But with HIV, some of it also reflects a shift from the more costly late-stage vaccine and microbicide trials that haven't worked. We are moving back to basic research, which is cheaper."

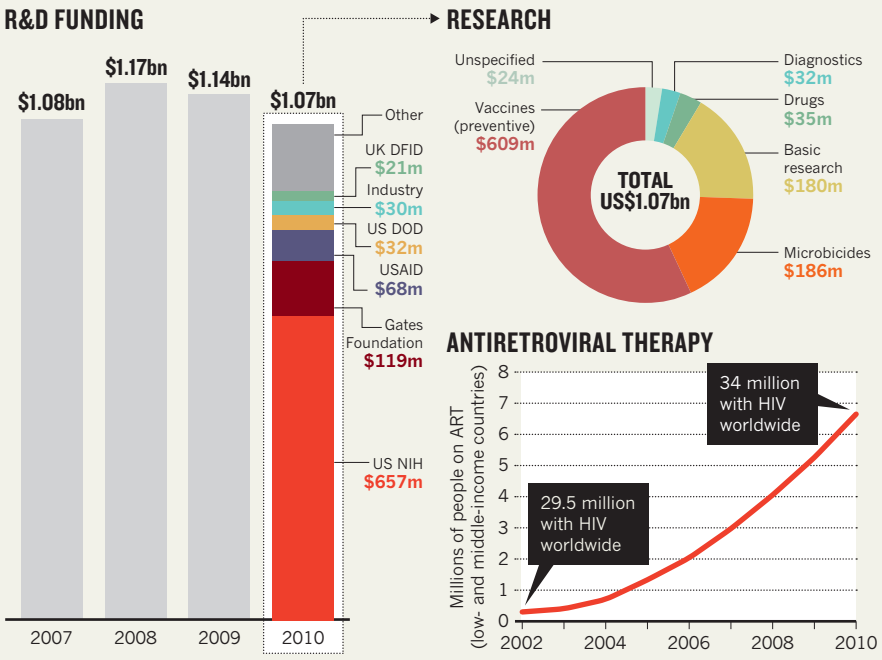
The cutbacks coincide with encouraging results on HIV/AIDS treatment. One study published in August found that early treatment of those infected greatly decreases sexual transmission to their uninfected partners, (M. S. Cohen *et al. N. Engl. J. Med.* **365,** 493–505; 2011). Other recent work suggests that ART, although costly, will pay for itself and more by increasing labour productivity, reducing the need for orphan care and deferring medical treatment costs (S. Resch *et al. PLoS ONE* **6,** e25310; 2011). "Not only is [treatment] the right thing to do, it's also the smart thing to do," says Sharonann Lynch, an HIV policy adviser with MSF in New York.

But the 'treatment equals prevention' equation invoked by Obama is not selling in all quarters. Roger Bate, a health economist and HIV expert at the American Enterprise Institute, a think tank in Washington DC, says that he is not yet persuaded by the studies suggesting that treatment prevents transmission. "There is a lot of effort out there to ignore noise in those studies," he says.

He adds that, even if treatment does prevent transmission, the commitment to treat infected people for their lifetimes quickly becomes an enormous financial outlay in future years. "Is it sustainable to treat even more people under current budgets?" he asks. "I don't think it is." ■

FIGHTING A KILLER

Funding for research and development (R&D) of HIV/AIDS products targeted at developing countries fell last year, whereas treatment with antiretroviral therapies has grown rapidly over the past decade.



SOURCE: POLICY CURES

Figures are adjusted for inflation and reported in 2007 US dollars. ART, antiretroviral therapy; DFID, Department for International Development; DOD, Department of Defense; NIH, National Institutes of Health; USAID, US Agency for International Development.

NANOTECHNOLOGY

Nano rules fall foul of data gap

Incomplete nanotoxicology research is hampering efforts to introduce regulation.

BY EUGENIE SAMUEL REICH

The science of the very small will lurk in many a Christmas gift this year. Nanomaterials add strength to golf clubs, odour resistance to socks and ultraviolet protection to cosmetics. But some of their properties could also pose health risks, and regulations covering their manufacture and use have failed to keep pace with the rush to market.

With sizes measured in mere billionths of a metre, nanomaterials have very high surface area relative to their mass, and this could alter their toxicity compared with the same material in bulk. Yet many nanomaterials are not regulated separately — in part because the high-quality nanotoxicity research needed for regulation is lacking, as an assessment presented at last week's Materials Research Society meeting in Boston, Massachusetts, made clear.

Groups representing food manufacturers had asked Cantox Health Sciences International, a consultancy in Mississauga, Canada, to review the literature for studies that could shed light on the effects of consuming nanoparticles from food packaging, in which they may be used to improve strength or even detect spoilage.

Cantox sampled 30 papers describing studies in which nanoparticles were fed to animals or applied *in vitro* to cells of the gastrointestinal tract. Of 21 animal studies, the Cantox researchers found that 15 failed to attain the highest grade on the Klimisch test, a quality-control criterion that assesses whether basic data, such as the number or gender of animals used, have been reported. "Some didn't give the dosage. This is a fairly important aspect," says Jeffrey Card, a toxicologist at Cantox who led the study.

The underlying materials science was equally spotty. Of 10 parameters that Cantox concluded should be reported in a nanotoxicity

© 2011 Macmillan Publishers Limited. All rights reserved

study, such as particle size, shape and surface chemistry, more than half were missing from 14 of the animal studies[1].

Yongping Bao of the University of East Anglia in Norwich, UK, an author on one of the studies[2] given low marks by Cantox, points out that some of the missing data appear in other papers by his group. But he believes that the review highlights the need for government regulatory agencies to establish agreed standards for this sort of research. "Anyone can set his or her own criteria to review nanotoxicity literature, so a gold standard is urgently needed," he says. Nanotoxicity studies arguably require additional standards, because nanoparticles can vary within a batch much more widely than do bulk chemicals, for example.

Altaf Carim, assistant director of nanotechnology at the White House Office of Science and Technology Policy, says that the US National Nanotechnology Initiative's Environmental, Health, and Safety Research Strategy[3], unveiled in October, will develop a set of standards that should help researchers and agencies to plug that gap. "We are certainly hoping the broader community is going to make use of it," Carim says.

Nanotoxicologists have also lacked standard reference materials for defining nanoparticles. The US National Institute of Standards and Technology has produced references for gold and polystyrene nanoparticles, but has yet to issue similar samples of other nanomaterials.

Later this month, however, the institute will release its first batch of standard reference carbon nanotubes — the result of three years' work, according to Kalman Migler, who is overseeing the project. Researchers will be able to buy vials of the nanotubes to calibrate their measurement equipment and assess levels of common contaminants in their samples. The reference nanotubes, the first ever provided by a national standards agency, should help nanotoxicologists to pinpoint whether any observed adverse reactions are the result of nanoparticles rather than impurities, for example.

Yet the latest nanotechnology decree from Europe perhaps shows how far regulation efforts still have to go. On 18 October, the European Commission finally adopted a definition of a nanomaterial: one in which half or more of the particles in a sample have one dimension that measures between 1 and 100 nanometres[4]. The definition had been requested by the European Parliament in April 2009. ■

1. Card, J. W., Jonaitis, T. S., Tafazoli, S. & Magnuson, B. A. *Crit. Rev. Toxicol.* **41,** 20–49 (2011).
2. Zhang, J.-S., Gao, X.-Y., Zhang, L.-D. & Bao, Y.-P. *BioFactors* **15,** 27–38 (2001).
3. *NNI Environmental, Health, and Safety Research Strategy* (US Natl Sci. Technol. Council, 2011); available at http://go.nature.com/8hkuv3
4. Potočnik, J. *Official J. EU* L275/38–L275/40 (2011).

PHARMACEUTICALS

Novartis to shut brain research facility

Drug giant redirects psychiatric efforts to genetics.

BY ALISON ABBOTT

Novartis has joined other pharmaceutical companies in abandoning traditional drug-discovery programmes that pursue treatments for brain disorders.

Nature has learned that the company is closing its neuroscience facility in Basel, Switzerland, where Novartis is headquartered.

It follows similar moves by GlaxoSmithKline and AstraZeneca, both based in the United Kingdom, which last year announced the closure of all their neuroscience research divisions globally. US-based companies Pfizer and Merck, as well as the French company Sanofi, have also pulled back on research into brain disorders. Rather than abandon neuroscience, however, Novartis says it will launch new programmes to study the genetics of psychiatric and cognitive disorders, in the hope of identifying new treatment strategies.

Developing drugs for brain disorders has become a high-risk activity, with most candidates failing after years of expensive clinical trials. The market is already flooded with cheap, generic antidepressants, antipsychotics and other drugs that act on known targets in the brain, mostly neurotransmitter receptors (see 'Mental-health drug sales'). This has forced companies to look for radically new drug targets, but the search is proving difficult because little is known about the biology of the brain and its disorders.

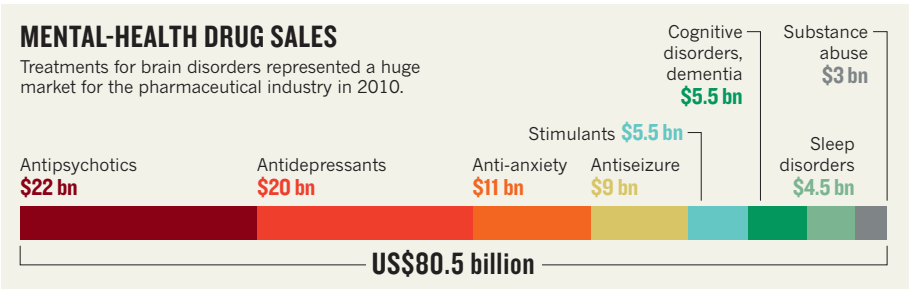
"Standard approaches to developing drugs for mental health have not reaped significant benefit in the past two decades," says Ken Kaitin, director of the Tufts Center for the Study of Drug Development in Boston, Massachusetts. "But it is a dilemma for the companies because there is a large and growing market for these products." Mental disorders impose the largest disease burden worldwide (see *Nature* **478,** 15; 2011), and current treatments do not work particularly well for most patients.

The Basel neuroscience department has 80 staff, and five candidate drugs in the pipeline that act on the central nervous system (CNS). These drugs will continue to be developed, says geneticist Mark Fishman, president of research for the company's main research arm, the Novartis Institutes of BioMedical Research (NIBR). Novartis is negotiating with unions over the terms of the closure, which is likely to happen during 2012.

Fishman says that the NIBR will open a new research division to study the genetics of psychiatry and cognitive disorders at its site in Cambridge, Massachusetts, taking advantage of academic strength in the area, including the Broad Institute, run jointly by Harvard University and the Massachusetts Institute of Technology, which has both major DNA-sequencing capacity and a large group dedicated to psychiatric genetics. "Progress based on neurotransmitters has become small and incremental," Fishman told *Nature*. "Genetic analysis will provide a real scientific opportunity in psychiatric and cognitive disorders, even if new drugs only arrive in the distant future."

"Developing drugs for the brain has become high-risk, with most candidates failing."

Recent population-based, genome-wide association studies have identified many different genetic variants that individually confer tiny increased risk of disease, but which may act together to substantially increase a person's susceptibility. A type of defect known as copy number variation, in which large chunks of DNA are gained or lost, may on its own cause a significant increase in risk for psychiatric disease. Scientists hope that understanding the ▶



© 2011 Macmillan Publishers Limited. All rights reserved

study, such as particle size, shape and surface chemistry, more than half were missing from 14 of the animal studies[1].

Yongping Bao of the University of East Anglia in Norwich, UK, an author on one of the studies[2] given low marks by Cantox, points out that some of the missing data appear in other papers by his group. But he believes that the review highlights the need for government regulatory agencies to establish agreed standards for this sort of research. "Anyone can set his or her own criteria to review nanotoxicity literature, so a gold standard is urgently needed," he says. Nanotoxicity studies arguably require additional standards, because nanoparticles can vary within a batch much more widely than do bulk chemicals, for example.

Altaf Carim, assistant director of nanotechnology at the White House Office of Science and Technology Policy, says that the US National Nanotechnology Initiative's Environmental, Health, and Safety Research Strategy[3], unveiled in October, will develop a set of standards that should help researchers and agencies to plug that gap. "We are certainly hoping the broader community is going to make use of it," Carim says.

Nanotoxicologists have also lacked standard reference materials for defining nanoparticles. The US National Institute of Standards and Technology has produced references for gold and polystyrene nanoparticles, but has yet to issue similar samples of other nanomaterials.

Later this month, however, the institute will release its first batch of standard reference carbon nanotubes — the result of three years' work, according to Kalman Migler, who is overseeing the project. Researchers will be able to buy vials of the nanotubes to calibrate their measurement equipment and assess levels of common contaminants in their samples. The reference nanotubes, the first ever provided by a national standards agency, should help nanotoxicologists to pinpoint whether any observed adverse reactions are the result of nanoparticles rather than impurities, for example.

Yet the latest nanotechnology decree from Europe perhaps shows how far regulation efforts still have to go. On 18 October, the European Commission finally adopted a definition of a nanomaterial: one in which half or more of the particles in a sample have one dimension that measures between 1 and 100 nanometres[4]. The definition had been requested by the European Parliament in April 2009. ■

1. Card, J. W., Jonaitis, T. S., Tafazoli, S. & Magnuson, B. A. *Crit. Rev. Toxicol.* **41,** 20–49 (2011).
2. Zhang, J.-S., Gao, X.-Y., Zhang, L.-D. & Bao, Y.-P. *BioFactors* **15,** 27–38 (2001).
3. *NNI Environmental, Health, and Safety Research Strategy* (US Natl Sci. Technol. Council, 2011); available at http://go.nature.com/8hkuv3
4. Potočnik, J. *Official J. EU* L275/38–L275/40 (2011).

PHARMACEUTICALS

Novartis to shut brain research facility

Drug giant redirects psychiatric efforts to genetics.

BY ALISON ABBOTT

Novartis has joined other pharmaceutical companies in abandoning traditional drug-discovery programmes that pursue treatments for brain disorders.

Nature has learned that the company is closing its neuroscience facility in Basel, Switzerland, where Novartis is headquartered.

It follows similar moves by GlaxoSmithKline and AstraZeneca, both based in the United Kingdom, which last year announced the closure of all their neuroscience research divisions globally. US-based companies Pfizer and Merck, as well as the French company Sanofi, have also pulled back on research into brain disorders. Rather than abandon neuroscience, however, Novartis says it will launch new programmes to study the genetics of psychiatric and cognitive disorders, in the hope of identifying new treatment strategies.

Developing drugs for brain disorders has become a high-risk activity, with most candidates failing after years of expensive clinical trials. The market is already flooded with cheap, generic antidepressants, antipsychotics and other drugs that act on known targets in the brain, mostly neurotransmitter receptors (see 'Mental-health drug sales'). This has forced companies to look for radically new drug targets, but the search is proving difficult because little is known about the biology of the brain and its disorders.

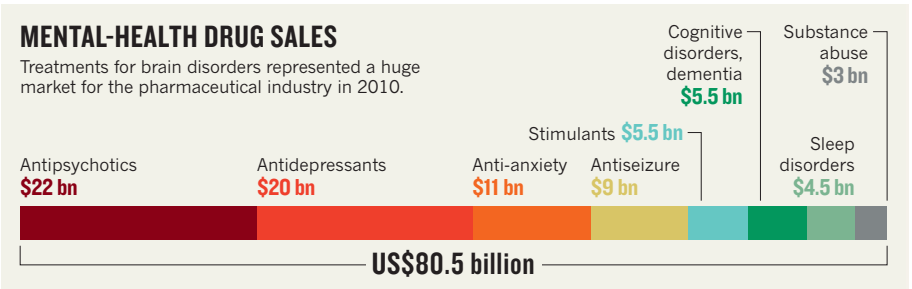
"Standard approaches to developing drugs for mental health have not reaped significant benefit in the past two decades," says Ken Kaitin, director of the Tufts Center for the Study of Drug Development in Boston, Massachusetts. "But it is a dilemma for the companies because there is a large and growing market for these products." Mental disorders impose the largest disease burden worldwide (see *Nature* **478,** 15; 2011), and current treatments do not work particularly well for most patients.

The Basel neuroscience department has 80 staff, and five candidate drugs in the pipeline that act on the central nervous system (CNS). These drugs will continue to be developed, says geneticist Mark Fishman, president of research for the company's main research arm, the Novartis Institutes of BioMedical Research (NIBR). Novartis is negotiating with unions over the terms of the closure, which is likely to happen during 2012.

Fishman says that the NIBR will open a new research division to study the genetics of psychiatry and cognitive disorders at its site in Cambridge, Massachusetts, taking advantage of academic strength in the area, including the Broad Institute, run jointly by Harvard University and the Massachusetts Institute of Technology, which has both major DNA-sequencing capacity and a large group dedicated to psychiatric genetics. "Progress based on neurotransmitters has become small and incremental," Fishman told *Nature*. "Genetic analysis will provide a real scientific opportunity in psychiatric and cognitive disorders, even if new drugs only arrive in the distant future."

"Developing drugs for the brain has become high-risk, with most candidates failing."

Recent population-based, genome-wide association studies have identified many different genetic variants that individually confer tiny increased risk of disease, but which may act together to substantially increase a person's susceptibility. A type of defect known as copy number variation, in which large chunks of DNA are gained or lost, may on its own cause a significant increase in risk for psychiatric disease. Scientists hope that understanding the ▶



© 2011 Macmillan Publishers Limited. All rights reserved

biological function of these risk genes may yield new clues about how to prevent or treat diseases.

Companies also hope that they will be able to find genetic biomarkers — a particular DNA sequence or variation, for example — that will allow them to identify individuals who are likely to respond to a particular drug, says Kaitin. "It's the basis of personalized medicine — health plans are more likely to pay for a new drug if they are convinced it will work for the person it is prescribed for."

Stephen Stahl, a psychopharmacologist and psychiatrist at the University of California, San Diego, says that other companies are also using gene sequencing to look for new targets. "GlaxoSmithKline pulled out of traditional CNS pharmacology but is pursuing research into neurodegeneration at its new sequencing centre in Shanghai." It's a promising approach, he adds, but he notes that no one yet knows how successful it will be.

The genetic approach has delivered encouraging results for a five-year, €20-million (US$27-million) public–private partnership called NEWMEDS, launched in 2009 by the European Commission to speed the development of treatments for schizophrenia and depression. NEWMEDS coordinator Tine Bryan Stensbøl, director of research at the Danish pharmaceutical company H. Lundbeck in Valby, says that scientists in the consortium have used genome information collected from the Icelandic population by the Reykjavik-based genetics company deCODE to identify a copy number variant that they believe will be important in understanding the pathology of schizophrenia. The commission has awarded NEWMEDS a €1-million top-up to investigate further.

Lundbeck is a medium-sized company that remains committed to drug discovery in psychiatric disease. It has, for example, just signed an agreement worth more than a billion dollars with the Japanese company Otsuka Pharmaceutical to develop and market products for mental disorders. "We are certain that the rewards will come and I'm glad we will be well-placed to enjoy them," says Stensbøl. But she regrets that other companies are moving away, as she believes that competition can stimulate the field: "We need a critical mass of companies there when the new targets come."

Johnson & Johnson is one of the few major international firms keeping the faith. Its head of neuroscience, Husseini Manji, who is based in New Jersey, says that "things are cyclical — for those who stay the course, the breakthroughs will finally come". ■

➲ NATURE.COM
See *Nature*'s schizophrenia special:
go.nature.com/ctwqay



S. TRIANA

Daniel Zimmerman is part of a team that has been working for ten years to simulate Earth's magnetic field.

EARTH SCIENCE

Dynamo maker ready to roll

Two rotating spheres separated by thousands of kilograms of liquid sodium aim to mimic Earth's interior.

BY SUSAN YOUNG

The 3-metre-tall ribbed sphere looks like a 'visitor' from another planet, dramatically lit and encased within a sturdy steel box. In fact, the giant orb, housed in a cavernous warehouse at the University of Maryland, College Park, is meant to approximate Earth's core.

Ten years in the making, the US$2-million project is nearly ready for its inaugural run. Early next year, the sphere will begin whirling around while loaded with 13,000 kilograms of molten sodium heated to around 105 °C. Researchers hope that the churning, electrically conducting fluid will generate a self-sustaining electromagnetic field that can be poked, prodded and coaxed for clues about Earth's dynamo, which is generated by the movement of liquid iron in the outer core. If it works, it will be the first time that an

© 2011 Macmillan Publishers Limited. All rights reserved

biological function of these risk genes may yield new clues about how to prevent or treat diseases.

Companies also hope that they will be able to find genetic biomarkers — a particular DNA sequence or variation, for example — that will allow them to identify individuals who are likely to respond to a particular drug, says Kaitin. "It's the basis of personalized medicine — health plans are more likely to pay for a new drug if they are convinced it will work for the person it is prescribed for."

Stephen Stahl, a psychopharmacologist and psychiatrist at the University of California, San Diego, says that other companies are also using gene sequencing to look for new targets. "GlaxoSmithKline pulled out of traditional CNS pharmacology but is pursuing research into neurodegeneration at its new sequencing centre in Shanghai." It's a promising approach, he adds, but he notes that no one yet knows how successful it will be.

The genetic approach has delivered encouraging results for a five-year, €20-million (US$27-million) public–private partnership called NEWMEDS, launched in 2009 by the European Commission to speed the development of treatments for schizophrenia and depression. NEWMEDS coordinator Tine Bryan Stensbøl, director of research at the Danish pharmaceutical company H. Lundbeck in Valby, says that scientists in the consortium have used genome information collected from the Icelandic population by the Reykjavik-based genetics company deCODE to identify a copy number variant that they believe will be important in understanding the pathology of schizophrenia. The commission has awarded NEWMEDS a €1-million top-up to investigate further.

Lundbeck is a medium-sized company that remains committed to drug discovery in psychiatric disease. It has, for example, just signed an agreement worth more than a billion dollars with the Japanese company Otsuka Pharmaceutical to develop and market products for mental disorders. "We are certain that the rewards will come and I'm glad we will be well-placed to enjoy them," says Stensbøl. But she regrets that other companies are moving away, as she believes that competition can stimulate the field: "We need a critical mass of companies there when the new targets come."

Johnson & Johnson is one of the few major international firms keeping the faith. Its head of neuroscience, Husseini Manji, who is based in New Jersey, says that "things are cyclical — for those who stay the course, the breakthroughs will finally come". ■

➲ NATURE.COM
See *Nature*'s schizophrenia special:
go.nature.com/ctwqay



S. TRIANA

Daniel Zimmerman is part of a team that has been working for ten years to simulate Earth's magnetic field.

EARTH SCIENCE

Dynamo maker ready to roll

Two rotating spheres separated by thousands of kilograms of liquid sodium aim to mimic Earth's interior.

BY SUSAN YOUNG

The 3-metre-tall ribbed sphere looks like a 'visitor' from another planet, dramatically lit and encased within a sturdy steel box. In fact, the giant orb, housed in a cavernous warehouse at the University of Maryland, College Park, is meant to approximate Earth's core.

Ten years in the making, the US$2-million project is nearly ready for its inaugural run. Early next year, the sphere will begin whirling around while loaded with 13,000 kilograms of molten sodium heated to around 105 °C. Researchers hope that the churning, electrically conducting fluid will generate a self-sustaining electromagnetic field that can be poked, prodded and coaxed for clues about Earth's dynamo, which is generated by the movement of liquid iron in the outer core. If it works, it will be the first time that an

© 2011 Macmillan Publishers Limited. All rights reserved

experiment that mirrors the configuration of Earth's interior has managed to recreate such a phenomenon.

"Dynamos are easy to generate in nature," says project leader and experimental geophysicist Daniel Lathrop. "The same is not true for the lab." Yet simulating this process in the laboratory will guide our understanding of processes that take place 3,000 kilometres below ground, at depths that can be probed only indirectly — by analysing seismic waves that have travelled through Earth's deep interior, for example.

"There's no way for us to measure anything near the core directly," says Lathrop. "It exceeds the deepest well by a factor of 100."

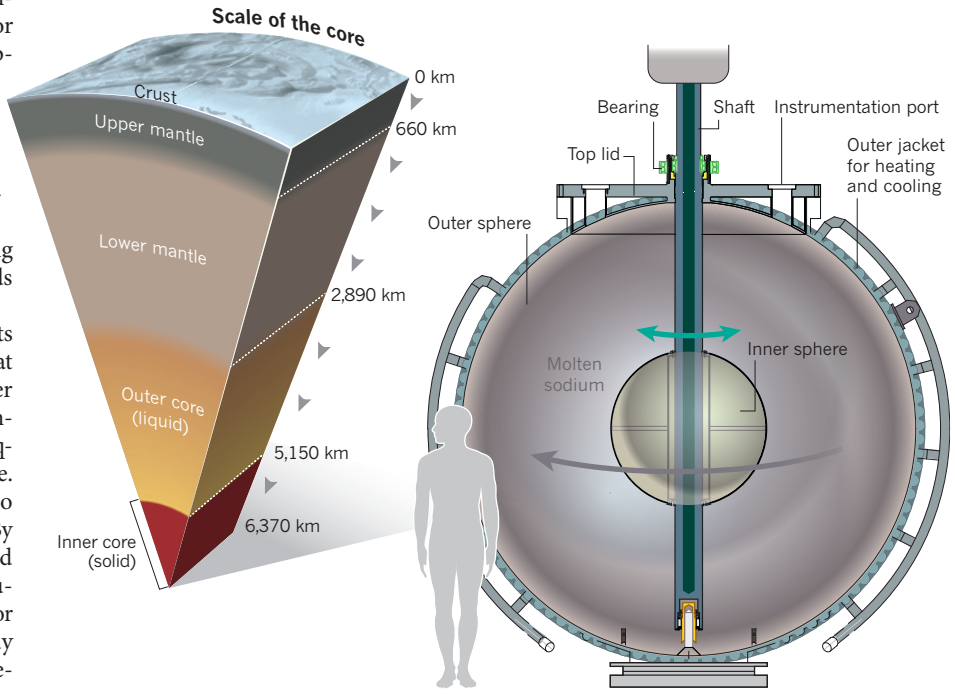
The University of Maryland set-up consists of two concentric spheres. The inner sphere, at 1 metre across, stands in for Earth's solid inner core; the outer sphere the edge of Earth's mantle. The space between the two is filled with liquid sodium, mimicking the liquid outer core. Each sphere is powered by a separate motor so that it can rotate independently of the other. By spinning the spheres across a range of matched and unmatched velocities — up to 4 revolutions per second for the outer sphere and 12 for the inner — Lathrop and his team will study how heat and rotation might affect the movement of the molten iron in Earth's core.

But the question that looms largest is whether the churning sodium will mimic the geodynamo and generate a self-perpetuating magnetic field. The experiment will use Earth's natural magnetism as a 'seed field' to kick-start the process. As this field is dragged and stretched by the spinning, conducting liquid it will generate electric currents. Those currents will then create additional magnetic fields that, when sufficiently twisted around, can amplify themselves and drive the process forward. No one knows whether this feedback loop will work, says Lathrop, because "there are neither theory nor experiments at these parameters".

Enthusiasm for the effort is building beyond Lathrop's group. "Everyone in the community is waiting with bated breath," says Andrew Jackson, a geophysicist at the Swiss Federal Institute of Technology Zurich. "He's asking questions that we don't know the answer to."

RECREATING EARTH'S CORE IN THE LAB

Earth's core is inaccessible, but a sphere filled with liquid sodium at the University of Maryland may uncover some of its hidden properties. An instrumentation port in the rotating lid allows researchers to monitor temperature, fluid flow and other properties while the inner and outer spheres spin independently of each other.



NEAR EARTH

The Maryland experiment expands on earlier efforts such as that built by a collaboration between the École Normale Supérieure institutes in Paris and Lyon and the French Atomic Energy Commission in Saclay. Built in Cadarache, France, the experiment uses spinning iron discs to force turbulence in liquid sodium within a cylindrical case. In 2007, the team reported that the combination of motion and a seed magnetic field led to the formation of a dynamo with some resemblance to Earth's (M. Berhanu *et al. Europhys. Lett.* **77,** 59001; 2007). The next step is to explore dynamos in set-ups that more closely match reality, says Jean-François Pinton, an investigator with the French experiment, which is still running.

"Our experiment is a physicist's experiment," Pinton says, because it aims only to generate a dynamo, regardless of how realistic the set-up is. "Dan is after what Earth does."

Other groups have also approximated planetary shapes in liquid-sodium experiments, but they haven't managed to generate self-sustaining dynamos. A group at the University of Wisconsin–Madison has a 1-metre-diameter spherical experiment that doesn't rotate, but instead drives turbulence with internal propellers. A group in Grenoble, France, built a 40-centimetre rotating set-up with a sodium-filled outer sphere and a solid inner sphere. Both experiments have uncovered interesting properties of fluid dynamics of conducting liquids, but not created a dynamo.

"Dynamos are easy to generate in nature. The same is not true for the lab."

Lathrop's latest construction is the fourth and largest in a series of sodium-spinners built by his group. None has yet generated a dynamo, but the size of the new sphere should give it a better chance of doing so.

Lathrop and others hope that their work will shed light on how rotational forces in Earth's core deflect flows of electrically conducting liquid into a configuration that produces a magnetic field with north and south poles. It might also help to explain what triggers changes in Earth's magnetic field such as north–south pole reversals, which occur on average every few hundred thousand years, but can happen after a few tens of thousands of years or a few million. The last reversal was some 780,000 years ago. Over the past century and a half, the magnetic field has weakened by about 10%, leading some researchers to suggest that another pole-reversal is in the offing.

Although a pole reversal is unlikely to cause any major disruptions, the weakening of Earth's magnetic field is "at least mildly alarming", says Lathrop, because the field shields us from charged solar particles, which can disrupt power grids (see *Nature* **441,** 402–404; 2006). "Earth's magnetic shield is part of what makes this planet hospitable," says Lathrop.

The sphere is now nearly three-quarters full and is awaiting its final shipment of solid sodium, expected next month. Once the last barrel of sodium is put in, the team will reconfigure the experiment for spinning.

"We've taken all the steps that we can think of to try to get dynamo action, but we don't have a detailed prediction about whether or not those steps will be successful," says Daniel Zimmerman, a postdoc at Maryland who has been with the project since the beginning.

But Lathrop isn't worried, saying that a large rotating volume of a liquid conductor is certain to churn up some surprises. "With or without the dynamo action it's going to be interesting," he says. ▪

NATURE.COM
For a video of the experiment in action see:
go.nature.com/guofxe

© 2011 Macmillan Publishers Limited. All rights reserved



REED SAXON/AP PHOTO

Student protesters disrupted board meetings at the University of California in November.

EDUCATION

State cuts fuel California protests

Fears raised of lasting damage to public university system.

BY ERIKA CHECK HAYDEN

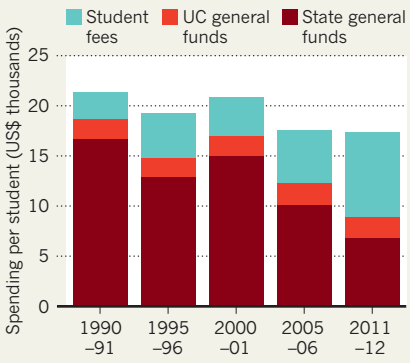
Reduced services. Shrinking programmes. Rising tuition fees. These are the signs of an ever more dire budget situation at the University of California, which has suffered years of state funding cutbacks.

Discomfort spread to the boardroom on 28 November, when student protesters at four campuses temporarily shut down a teleconference of the University of California regents, the body that sets policy and budgets for the ten-campus statewide system. The protests were sparked in part by an incident on 18 November, in which a policeman at the Davis campus doused student protesters in pepper spray, but the latest grievances also included ongoing hikes in undergraduate tuition fees, which have risen by 71% since 2008, and budget cuts that could reach US$750 million this year.

One of the world's leading public universities, the University of California system is reeling from the impact of the global financial crisis, decades of declining public support for education and a state budget crisis, says John Douglass, a higher-education researcher at the University of California, Berkeley. "The great recession has been a huge blow after a million small cuts, and now you're seeing a transformation of the funding model," he adds. The system increasingly relies on tuition fees, says Douglass. This year, for the first time, the University of California drew more of its funding from private sources, including fees, than from the state (see 'Getting less for more'). California now provides just 11% of the university's $22.5-billion budget, down from 30% in the 1980s.

GETTING LESS FOR MORE

Despite tuition-fee increases, spending per student at the University of California (UC) is now about 80% of what it was in 1990, owing to falling state funding.



Undergraduates have borne the brunt, facing higher fees, fewer courses and larger class sizes. Scientific researchers are less directly affected because they get much of their funding from federal grants and private foundations. But they are also feeling the effects of the squeeze as graduate students face constraints on resources and time, and campus services and maintenance are cut.

Philippe Marchand, a graduate student in environmental science and policy management at Berkeley, needs to increase the amount of paid teaching he does to cover his tuition fees. But teaching budgets across the system have fallen, so fewer teaching slots are available, and he is unsure whether the option will be available next semester. But even if he can teach, it will not solve his problems. "I'm going to have to make some difficult choices; if I have to teach every semester I may have to suspend my research," he says. Since 2006, the number of graduate students in Marchand's department has shrunk by 30%.

Jessica Smith, a chemistry graduate student at Berkeley, says that ceiling leaks in labs have ruined experiments and machinery, although one leak was fixed by funnelling the water away using rubber tubing. "The biggest budget-related impediment to my research is our crumbling facilities," she says.

The cuts are also making it harder for the university to retain staff, says Jeffrey Bluestone, executive vice-chancellor of the University of California, San Francisco (UCSF), which relies on state contributions to pay faculty salaries and benefits, and so cannot match the packages at some private institutions. Losses include Didier Stainier, a geneticist who went to the Max Planck Institute for Heart and Lung Research in Bad Nauheim, Germany, and Christopher Voigt, a synthetic biologist who left UCSF this year for the Massachusetts Institute of Technology in Cambridge. "In a different world, we could have kept him," Bluestone says of Voigt.

More broadly, ongoing cuts will hamper the system's ability to nurture California's technology sector, "which is heavily reliant on bringing in talent from universities", says Douglass.

A report on the future of the University of California, released in November 2010, led campuses to increase enrolment of students from other states, who pay higher fees. Another fund-raising option, not yet adopted, is to allow the most prestigious campuses, such as Berkeley and Los Angeles, to charge higher tuition fees than the others. And in state elections next year, voters will probably be asked to consider boosting taxes, in part to fund higher education.

Douglass says that such measures are unlikely to reverse the major declines in state funding. The University of California must try to staunch the bleeding before its reputation as a successful alternative to costly private US universities is irreparably damaged. ■

© 2011 Macmillan Publishers Limited. All rights reserved

OCEANOGRAPHY

China pushes to rule the waves

Launch of huge Kexue *research vessel could put country at forefront of ocean science.*

BY JANE QIU IN BEIJING

The most sophisticated ship yet to join China's research fleet was launched last week, marking the latest step in the country's push to boost marine science.

Dubbed *Kexue* (which means 'science'), the 100-metre-long vessel dwarfs research ships operated by some other oceanography powerhouses, such as the United States. It "will be one of China's key research vessels in the next couple of decades", said Ding Zhongli, vice-president of the Chinese Academy of Sciences (CAS), at the launch ceremony on 30 November. "It is a state-of-the-art moving laboratory for fundamental research and technological development in marine science, especially deep-sea research."

Kexue, which cost about 550 million renminbi (US$86 million), will be able to carry 80 people at a top speed of 15 knots on voyages of 28,000 kilometres, lasting up to 60 days. It is modelled on the United Kingdom's RRS *James Cook* and Norway's *G.O. Sars*, says Sun Song, director of the CAS Institute of Oceanology in Qingdao, Shandong province, which is in charge of *Kexue*'s design and operation. "It is on a par with these top research vessels," he says.

Most ocean-going research vessels in China have been designed to explore for mineral or fuel resources, and although CAS and some universities have more than half a dozen vessels fitted out for scientific studies, few have the capacity to explore much further than China's coastal waters. "The lab space on those ships is very limited and facilities are not up to scratch," says Zhao Meixun, a marine ecologist at the Ocean University of China (OUC) in Qingdao. For example, *Dongfanghong 2*, an OUC research vessel, lacks the relevant sampling and measurement tool to study sea water's iron content — a vital nutrient for algae that absorb carbon dioxide. It also lacks equipment for detecting trace elements, or carrying out complex biological studies.

With 360 square metres of lab space, *Kexue* can accommodate a wide range of oceanographic studies, including measurements of ocean and atmospheric circulation, water and sediment sampling and on-board analyses, deep-sea exploration, acoustic studies and ocean-floor seismic experiments. It can also launch manned submersibles, such as the *Jiaolong* ('Sea dragon') craft, which descended to a depth of 5,000 metres in July.



XINHUA/PHOTOSHOT

The *Kexue* research vessel was launched in Wuhan, China.

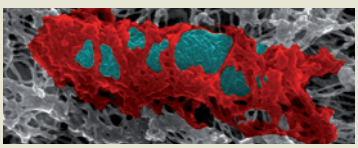
"*Kexue* is an impressive ship with a lot of cool features," says Shangping Xie, a climatologist and oceanographer at the University of Hawaii in Honolulu. Its propellers, driven by electric motors, are housed inside pods that can be rotated in any horizontal direction, making the ship much more manoeuvrable than if it had a fixed propeller and rudder system, says Sun. The electric motors, together with noise- and vibration-reduction measures, make the ship particularly quiet — ideal for marine-biology research. *Kexue* is the first research vessel in the world to boast these sophisticated propulsion pods, which together with positioning and motion-reduction systems will allow it to maintain a fixed position in high winds and rough seas, says Sun.

Yet the ship could prove too costly for some researchers. Unlike countries, such as Japan and the United States, that bear the costs of scientists' ship time, China will ask scientists to pay a substantial portion of the costs of operating *Kexue* out of their own grants, says Xie. Weijun Cai, an oceanographer at the University of Georgia in Athens, says that *Kexue* should be complemented with more small- and medium-sized ships that allow small teams of scientists to carry out focused research on a lower budget. "This is particularly important for the majority of young and less well-funded researchers," he says.

Other researchers say that to realize *Kexue*'s full potential, there should be more sharing of data and equipment between marine scientists, and better technical training of on-board staff to ensure that data sets from different missions are comparable.

On *Kexue*'s maiden voyage to the western Pacific Ocean next summer, scientists on board will study marine biodiversity, the deep-sea environment and the ocean's effect on climate in China. Researchers around the world are welcome to apply for ship time, says Sun. "*Kexue* will be a platform for joint marine research and observation with our foreign colleagues in the coming decades." ■

MORE ONLINE



S. MOSTOWY

TOP STORY

Septin proteins take bacterial prisoners go.nature.com/ri7s41

MORE NEWS

- Entangled diamonds vibrate together go.nature.com/rgfuf3
- Independent study affirms three-quarters of climate change is man-made go.nature.com/lhjft3
- Voyagers detect birth pains of stars go.nature.com/bbrxw2

VIDEO



A. THURBER

Yeti crabs living near methane seeps cultivate gardens of bacteria on their claws for food go.nature.com/ardltw

© 2011 Macmillan Publishers Limited. All rights reserved



D. DOUBILET/NATL. GEOGR.

Uncertain sanctuary

Nations are racing to establish marine protected areas, but it's not clear whether many are living up to the name.

BY DANIEL CRESSEY

Former US President George W. Bush did not garner much applause from environmentalists during his eight years in the White House, but on 15 June 2006, he gave them something to cheer about. Bush signed an order to create the Papahānaumokuākea Marine National Monument in Hawaii, then the world's largest ocean conservation area. Spanning about 360,000 square kilometres of the Pacific Ocean, the reserve is designed to safeguard 7,000 marine species including the rare Hawaiian monk seal (*Monachus schauinslandi*), hawksbill (*Eretmochelys imbricata*) and leatherback turtles (*Dermochelys coriacea*) and nearly two dozen other animals on the US endangered species list. It also started a race between other nations to follow suit with giant reserves of their own.

Since then, Australia, the United Kingdom and the Pacific nation of Kiribati have created still larger ocean conservation zones. And nations are discussing plans to make other huge reserves that will encompass much of the Southern Ocean and the Sargasso Sea. Worldwide, about 4 million square kilometres have now been set aside in marine protected areas (MPAs).

Despite that rapid expansion, many of the MPAs fall well short of the grand plans that governments once trumpeted to protect the world's oceans and the species that live in them. The 2002 Convention on Biological Diversity called on the world to protect 10% of the oceans by 2012, but the MPAs created so far cover little more than 1%.

Australia plans to put protections in place for the Coral Sea.

Studies show that some of the MPAs, such as Australia's Great Barrier Reef Marine Park, have provided strong protection for crucial ecosystems[1]. But scientists fret that many MPAs exist only on paper and offer no real benefits to the ecosystems or the people within or near them. Some have design flaws, such as protecting the wrong habitats. In others, there are no efforts to enforce limits on fishing or other activities within the protected areas. "The biggest thing we can do wrong is to claim to be protecting, when in fact the design of the place or the nature of the management is such that there is no effective protection," says Peter Sale, a marine scientist and professor emeritus at Canada's University of Windsor in Port Carling, Ontario.

These problems are likely to grow as countries increasingly join forces to develop larger and more expensive international protected areas on the high seas. Enforcement and monitoring will be more difficult in these places than they are in coastal waters. Some scientists worry that these concerns are not being openly discussed. "There's almost a religion, which is nurtured by non-governmental organizations and some members of the science community, to the extent that critical examination of MPAs is exceedingly difficult to get published," says Sale.

A SEA CHANGE

The term MPA can cover a vast range of entities, from tiny reserves less than a square kilometre in size to Britain's 544,000-square-kilometre Chagos Marine Protected Area, which wraps around 55 small islands in the Indian Ocean (see 'Fencing the sea'). 'No-take' MPAs ban fishing outright, but others permit certain types of fishing in specific zones.

One of the forces driving the expansion of larger MPAs has been the Global Oceans Legacy programme of the Pew Environment Group in Washington DC. In the mid-2000s, Pew identified a handful of sites as suitable for mega-MPAs because they had relatively low economic value and were owned by countries that had a history of conservation and the ability to enforce protections.

© 2011 Macmillan Publishers Limited. All rights reserved

The United States has established MPAs in two of these areas: the Papahānaumokuākea protected zone and one in the Mariana Trench in the western Pacific. In late November, Australia proposed plans for a 990,000-square-kilometre reserve in the Coral Sea off the country's northeast coast. Another of the sites highlighted by Pew was New Zealand's Kermadec Islands, where protections are currently in place for the sea floor, and conservationists would like to extend them to the entire water column.

Jay Nelson, director of the global ocean legacy programme at Pew, sees the Papahānaumokuākea MPA as the marine equivalent of Yellowstone National Park, the first big parcel of land to be set aside as a reserve. The creation of that park in 1872, he says, set off a chain reaction, prompting other countries to establish large parks of their own. "That had a similar effect to what Papahānaumokuākea has had in the oceans."

Countries have picked up the pace of establishing reserves as the deadline approaches for the 2012 target. In 2008, Daniel Pauly, a fisheries biologist at the University of British Columbia in Vancouver, Canada, estimated that 2.35 million square kilometres of ocean had been protected[2]. In late 2010, a report from the United Nations Environment Programme (UNEP) and the International Union for Conservation of Nature and Nature Conservancy charities put the area at 4 million square kilometres, with the number of MPAs nearing 5,900.

That area represents just 1.17% of the world's seas. So the deadline for reaching 10% was extended to 2020 at last year's meeting of countries that have signed up to the UN Convention on Biological Diversity.

The recent spike in protected areas has Pauly feeling hopeful, despite the failure to meet the 2012 target. "Maybe vast areas of the ocean will be put under protection in time to prevent bad things from happening," he says.

RESERVES WORK

Scientists and conservationists agree that a good MPA can make a difference. In 2009, Sarah Lester, a marine ecologist at the University of California, Santa Barbara, and her colleagues examined published peer-reviewed reports on 124 no-take marine reserves in 29 countries[3]. They found documented increases in characteristics such as biomass, species richness and population size within the boundaries of the reserves.

"By and large, everyone agrees that MPAs can achieve their conservation goals if they are enforced," says Benjamin Halpern, a co-author on the study and director of the Center for Marine Assessment and Planning at Santa Barbara.

But many MPAs are worth little more than the paper they are written on because nothing is done to enforce the rules or monitor the area. Earlier this year, Peter Jones, an MPA researcher at University College London, and his colleagues produced a report[1] for UNEP analysing the governance of 20 MPAs, including those in iconic regions such as the Great Barrier Reef and the Galapagos. The team found that none of the MPAs rated more than three out of five points for their effectiveness at meeting key objectives, such as preserving biodiversity and fostering sustainable use of resources. "Three of the 20 MPAs we've said had no effectiveness whatsoever — completely paper parks," Jones says.

Of these, Baleie Franca in Brazil and the Cres-Lošinj reserve in Croatia scored one. Pirajubaé in Brazil scored zero, in part because the government approved a highway in the area that harmed coastal habitats and fishing grounds, the report says.

Even countries that do try to regulate their MPAs can find it hard to fund them. "We don't have the resources that we need to actually monitor, enforce and understand these areas," says Jane Lubchenco, head of the US National Oceanic and Atmospheric Administration.

Poor planning can also hamper an MPA. Mexico, for example, set aside parts of the Gulf of California to protect the endangered vaquita — *Phocoena sinus*, a type of porpoise — but did not include a large part of the species' range.

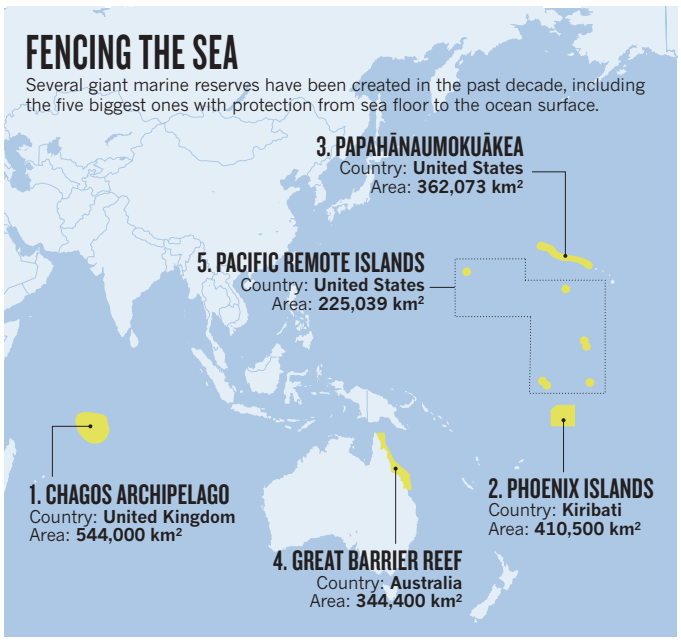
The Australian government has also been criticized over its efforts at marine zoning. In August, a group of 173 scientists published an open letter warning that a proposed 320,000-square-kilometre reserve off the southwestern coast "fails on the most basic test" by not protecting a representative selection of habitats in the region. Some suggest that Australia's proposal for a Coral Sea MPA faces similar issues (see *Nature* **480,** 14–15; 2011) .

In an e-mail to *Nature* about the southwestern reserve, a spokesperson from the Australian Department of Sustainability, Environment, Water, Population and Communities said that the government has a responsibility to balance social and economic factors along with ecological ones in decisions about conservation. "The department does not accept that there is a lack of protection for valuable habitat in the proposed marine reserves network," said the spokesperson.

Researchers say that there are ways to improve reserve planning. Many nations are now setting up MPAs using a software called Marxan, which uses an algorithm to calculate the areas that should be included in order to protect the most important habitat while minimizing the cost of the reserve. One of the driving forces behind the software, Hugh Possingham, head of the Spatial Ecology Lab at the University of Queensland in Brisbane, Australia, says that the best approach basically amounts to saying "get me 10% of everything and annoy as few people as possible". Australia used Marxan to design the Great Barrier Reef MPA, which sets aside large fractions of some 70 types of habitat.

Looking forward, researchers want to investigate how best to monitor and manage the increasing numbers of MPAs. As they assess the effectiveness of reserves, scientists sometimes find surprises, such as the fate of the monk seals in Hawaii's Papahānaumokuākea MPA. The seal population has increased outside the protected region but declined inside it, perhaps because sharks have thrived there[4]. Ashley McCrea Strub, a postdoctoral fellow working with Pauly's lab, says that there are very few data on the cost effectiveness of reserves and it will be vital to fill that gap. "That's really the future of where we need to go," she says. "That's an enormously important question."

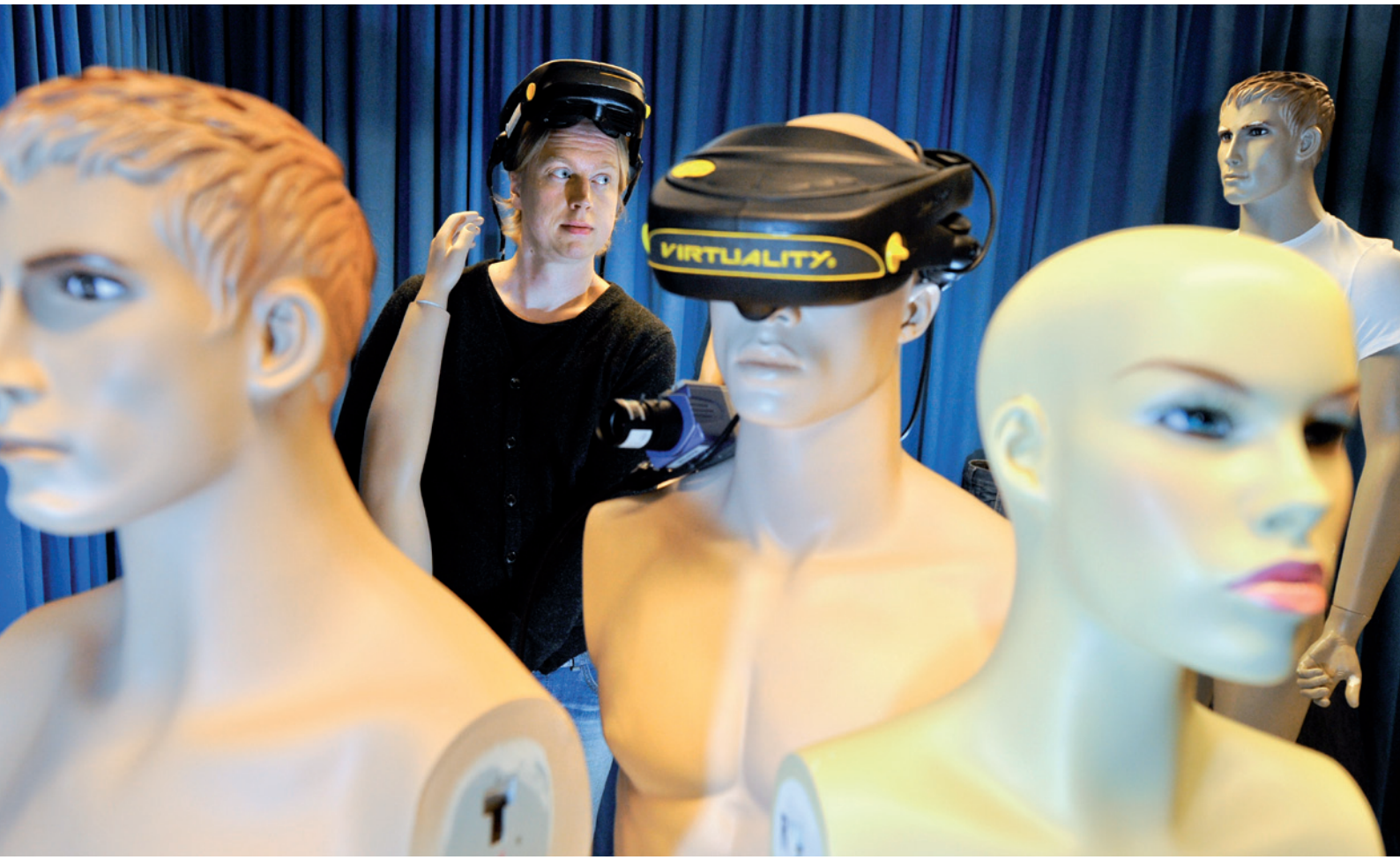
Some scientists have argued for doing more studies on MPAs before moving forward with major plans for new ones. But Possingham says the benefits in most cases are clear enough. "If you wait till you have perfect data," he says, "then you'll be waiting forever." ■ SEE EDITORIAL P.151



Daniel Cressey *is a reporter with* Nature *in London.*

1. Jones, P. J. S., Qiu, W. & De Santo, E. *Governing Marine Protected Areas: Getting the Balance Right* (UNEP, 2011); available at http://go.nature.com/tjyfbw
2. Wood, L. J., Fish, L., Laughren, J. & Pauly, D. *Oryx* **42,** 340–351 (2008).
3. Lester, S. E. *et al. Mar. Ecol. Prog. Ser.* **384,** 33–46 (2009).
4. Jones, N. *Nature* http://dx.doi.org/10.1038/news.2011.522 (2011).

© 2011 Macmillan Publishers Limited. All rights reserved



Master of illusion

Henrik Ehrsson uses mannequins, rubber arms and virtual reality to create body illusions, all in the name of neuroscience.

BY ED YONG

N. LARSSON/AP FOR NATURE

It is not every day that you are separated from your body and then stabbed in the chest with a kitchen knife.

But such experiences are routine in the lab of Henrik Ehrsson, a neuroscientist at the Karolinska Institute in Stockholm, who uses illusions to probe, stretch and displace people's sense of self. Today, using little more than a video camera, goggles and two sticks, he has convinced me that I am floating a few metres behind my own body. As I see a knife plunging towards my virtual chest, I flinch. Two electrodes on my fingers record the sweat that automatically erupts on my skin, and a nearby laptop plots my spiking fear on a graph.

Out-of-body experiences are just part of Ehrsson's repertoire. He has convinced people that they have swapped bodies with another person[1], gained a third arm[2], shrunk to the size of a doll or grown to giant proportions[3]. The storeroom in his lab is stuffed with mannequins of various sizes, disembodied dolls' heads, fake hands, cameras, knives and hammers. It looks like a serial killer's basement. "The other neuroscientists think we're a little crazy," Ehrsson admits.

But Ehrsson's unorthodox apparatus amount to more than cheap trickery. They are part of his quest to understand how people come to experience a sense of self, located within their own bodies. The feeling of body ownership is so ingrained that few people ever think about it — and those scientists and philosophers who do have assumed that it was unassailable.

"Descartes said that if there's something you can be certain of in this world, it's that your hand is your hand," says Ehrsson. Yet Ehrsson's illusions have shown that such certainties, built on a lifetime of experience, can be disrupted with just ten seconds of visual and tactile deception. This surprising malleability suggests that the brain continuously constructs its feeling of body ownership using information from the senses — a finding that has earned Ehrsson publications in *Science* and other top journals, along with the attention of other neuroscientists.

"A lot of people thought the sense of self

© 2011 Macmillan Publishers Limited. All rights reserved

was hard-wired, but it's not at all. It can be changed very quickly, and that's very intriguing," says Miguel Nicolelis, a neurobiologist at Duke University Medical Center in Durham, North Carolina.

Ehrsson's work also intrigues neuroscientists and philosophers because it turns a slippery, metaphysical construct — the self — into something that scientists can dissect. "We can say if we wobble the signals this way, our conscious experience wobbles in this way," says David Eagleman, a neuroscientist who studies perception at Baylor College of Medicine in Houston, Texas. "That's a lever we didn't have before."

"There are things like selfhood that people think cannot be touched by the hard sciences," says Thomas Metzinger, director of the Theoretical Philosophy Group at Johannes Gutenberg University of Mainz, Germany. "They are now demonstrably tractable. That's what I think is valuable about Henrik's contribution."

DAYDREAM BELIEVER

Ehrsson was born in 1972 in a suburb of Stockholm. His father was a chemist, his grandfather a dentist, and both stoked an interest in science and the human body that led him to study medicine at the Karolinska Institute. But in long anatomy classes, Ehrsson often found himself bored. "During lectures, I kept on thinking that if my eyes were floating over there and I was looking at myself from that perspective, where would my consciousness be?" After a pause: "I was not an A student."

After graduating, Ehrsson left medicine to start a PhD at the Karolinska, using brain scanners to study how people grasp objects. At the same time, he was becoming deeply fascinated by physical illusions. Some of these are well established: Aristotle discovered that crossing the index and middle fingers and touching the nose can — in some people — create a feeling of having two noses. Ehrsson also heard about the rubber-hand illusion, a trick devised in the late 1990s by researchers in the United States. They could convince people that a fake hand was their own by hiding their real hand under a table, placing a rubber one in front of them, and stroking both in the same way[4]. "It really worked on me," Ehrsson says. "It was fantastic and surreal."

Ehrsson started investigating illusions on the side. By the time he completed his postdoc at University College London and returned to the Karolinska Institute to start his own lab, illusions had become his main focus. He knew that many scientists use visual illusions to glean the basics of perception. "There are conferences on visual illusions, and yearly contests to create the best new ones," he says. "But there aren't many body illusions. The body has never been a main focus of psychology." It was these tricks, such as the rubber-hand illusion, that Ehrsson was keen to explore. He wanted to test how easily the sense of body ownership could be distorted.

Ehrsson set about creating more illusions based on the same principles as the rubber hand one. Headsets, cameras or fake body parts fooled the eyes, and synchronous strokes and prods added a tactile clincher. In 2007, he reported that he had used such props to convince subjects that they had left their own bodies[5] — a stunt that attracted headlines around the world.

At the time, some scientists and members of the public were openly sceptical that the illusion really worked. But on a trip to Ehrsson's lab this September, I was convinced. The goggles I wore displayed the view from a camera pointing at my back (see 'Out-of-body experience'). Ehrsson tapped my chest with one plastic rod while using a second one to synchronously prod at the camera. I saw and felt my chest being prodded at the same time as I saw a picture of myself from behind. Within ten seconds, I felt as if I was being pulled out of my real body and was floating several feet behind it.

> *"Henrik's work speaks to the idea that there is no such thing as a soul or a self that's independent of the brain."*

A year after removing his subjects from their own bodies, Ehrsson learned how to trick them into acquiring new ones. This time, the volunteers' goggles showed them the view from a camera on the head of a mannequin looking at its own plastic torso. Simultaneously poking the arm or stomach of the mannequin and the volunteer a few times was enough to convince the subjects that they were the dummy. They could even stare at their old bodies from their new ones and shake hands with their old self, all without breaking the spell[1]. "It really is very intense and incredibly fast," says Mark Hallett, a neurologist from the National Institutes of Health in Bethesda, Maryland, who experienced it first hand.

In his latest trick, published in May[3], Ehrsson convinced people that they had jumped into a tiny Barbie doll. When he prodded the doll's legs, the volunteers thought they were being prodded by giant objects. And when Ehrsson tested the illusion on himself and a colleague touched his cheek, he says, he looked up and "felt as if I was back in my childhood and looking at my mother".

Not everyone succumbs. Ehrsson suspects that people who can expertly localize their limbs without sight, such as dancers or musicians, would be less susceptible than the students with whom he normally works. But typically, the illusions work for around four out of five people. Ehrsson confirms the effects by asking his volunteers about their experiences and by threatening their disembodied, shrunken or plastic forms with a knife. If the illusion has worked, then volunteers show a reflexive burst of nervous sweating (as I did). Even knowing the knife is coming is no defence: as Ehrsson puts it, the illusions are "cognitively impregnable".

Earlier this year, Ehrsson even tweaked the rubber-hand illusion and convinced people they owned a third hand[2]. "He has taken those basic ideas and investigated how far you can push them," says neuroscientist Matthew Botvinick, one of the creators of the rubber-hand illusion, at Princeton University in New Jersey. "He's shown how extreme and how malleable the body representation can be."

SELF DELUSION

Ehrsson's next challenge is to work out what these illusions reveal about the brain. According to textbook wisdom, people build up a perception of their bodies using 'proprioception' — signals from the skin, muscles and joints that indicate the relative position of body parts. But Ehrsson's illusions show that vision and touch are also a crucial part of the mix, and that the brain builds a sense of self by constantly compiling information from all these senses. Proprioception may be telling the brain that the body is seated in a chair, but the carefully timed vision and touch signals in Ehrsson's illusion convince the brain that it is somewhere else entirely.

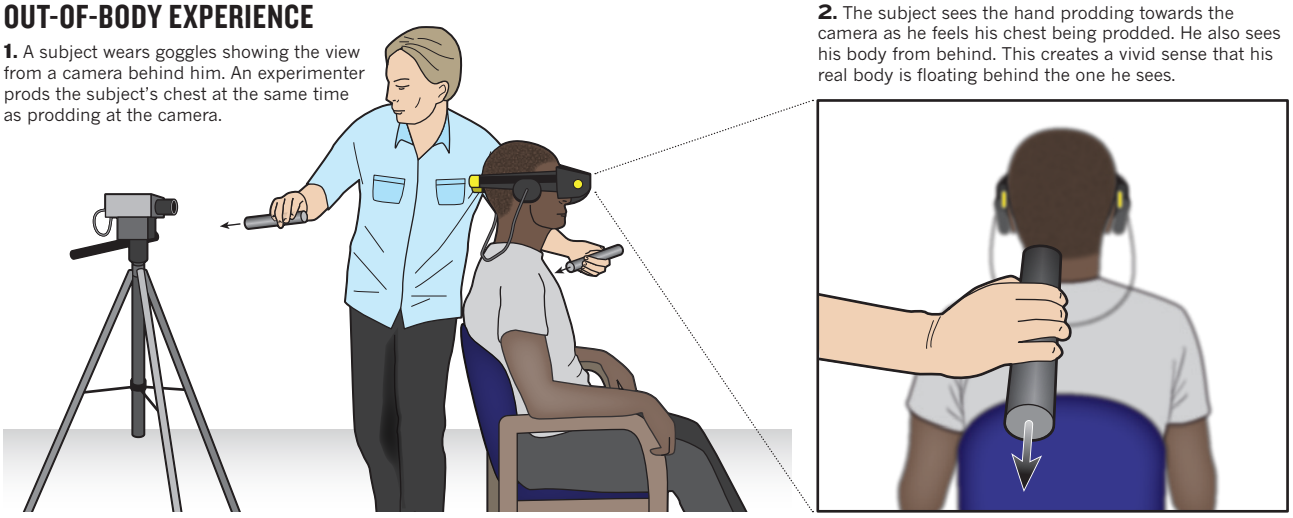
Ehrsson thinks such illusions depend on 'multisensory' neurons, which have been studied mostly in monkeys and allow the animals to successfully interact with objects by combining vision and touch. "We think these circuits are important, not just for representing external objects," says Ehrsson, "but representing your own body and the boundary between your body and the world." He thinks that the

© 2011 Macmillan Publishers Limited. All rights reserved

OUT-OF-BODY EXPERIENCE

1. A subject wears goggles showing the view from a camera behind him. An experimenter prods the subject's chest at the same time as prodding at the camera.

2. The subject sees the hand prodding towards the camera as he feels his chest being prodded. He also sees his body from behind. This creates a vivid sense that his real body is floating behind the one he sees.



neurons integrate information from across the senses to create a cohesive representation of the body. During an illusion, he simply changes the data flooding into the neurons to manipulate that representation.

For now, this is a working hypothesis. "The details of the multisensory integration have never been formally worked out," says Botvinick. "That's the missing link for me." Ehrsson and others are busy trying to see where these multisensory neurons reside in the human brain, by inducing illusions while volunteers sit inside a functional magnetic resonance imaging machine. The results differ. Ehrsson has found that the ventral premotor cortex — known to be involved in visual guidance of movements — is particularly active when people experience the full body-swap illusion[6]. Olaf Blanke at the University of Lausanne in Switzerland, one of only a handful of other researchers in this field, has shown that the nearby temporoparietal junction lights up when people experience out-of-body illusions[7]. He points out that brain damage or tumours in that area can induce disembodied feelings. "It's difficult to judge which is right, because we just have a thin neuroscientific layer of data at the moment," says Blanke.

SOUL SEARCHING

Ehrsson's interests in the self and body ownership bleed into the rest of his life. He is drawn towards experimental theatre and surrealist art. He comes across as thoughtful and remote, easily forgetting the names of colleagues and often lapsing into silence. He prefers working in his lab to going to conferences, he says.

He also occasionally gets angry letters from people who have had out-of-body experiences themselves. "They believe that their souls have left their bodies, and they feel threatened that a similar experience can be induced in a lab," says Ehrsson. He offers a diplomatic response, saying that he has "no way of disproving their ideas". Metzinger is more forthright. "Henrik's work speaks to the idea that there is no such thing as a soul or a self that's independent of the brain," he says.

Ehrsson is now working to turn his illusions towards a practical goal: developing better artificial limbs. Many people who have lost arms still feel the presence of their missing limbs, yet replacement hands feel alien. "We think that people might find it easier to use their prostheses, or develop a better body image, if we can elicit the illusion of real ownership," says Ehrsson.

To achieve this, he has adapted the rubber-hand illusion. With amputees, Ehrsson stimulates small spots on the stump that trigger the feeling of phantom fingers. By stroking these spots in time with the corresponding parts of a robotic 'cyberhand', Ehrsson and his colleagues have successfully convinced amputees that they owned the metal limb[8].

The effect disappeared about 10 to 15 seconds after the stroking stopped, though, so the stimulation would have to be continuous to keep the illusion going. That is what Ehrsson is working towards now. "The idea is to have an advanced prosthetic hand with sensors in the fingertips and stimulators in the stump," he says. Although other groups are working on similar sensory devices, Ehrsson says that his is different because he is attempting to precisely match up the feelings from the artificial fingers with stimulation in the corresponding part of the stump to generate the illusion of ownership.

Ehrsson has even bigger ambitions. Ownership illusions could help people to take control of entire alien bodies, both virtual and robotic, in a way that would afford a finer degree of control than the joysticks and other controllers used to steer robots and avatars today. People who control robots, for example, would see from their machine's perspective using goggles, steer using motion-capture suits and receive tactile feedback from gloves connected to sensors in the robot's hands. As long as signals are transferred between human and machine within 100 milliseconds, Ehrsson predicts that "the full-body illusion will kick in". And size should not matter. A surgeon could control a microscopic robot in a patient's body. Giant robots could repair broken oil rigs or fractured nuclear power stations. Ehrsson smiles as he considers the possibilities.

There is one illusion that even Ehrsson is unsure if he can pull off: splitting the sense of self, so that a person can commandeer two bodies. Ehrsson thinks that the trick's feasibility may depend on how the brain integrates information from the senses. "Maybe it's statistical in nature. Maybe instead of using absolute coordinates, the brain says it's most likely that my hand is here," he says. "If it works like that, maybe you could trick the brain into thinking that this body and that body are equally likely to be me."

It sounds far-fetched, but so does most of what Ehrsson has achieved so far. "We're working on it," he says. "Then again, it might be impossible." ■

NATURE.COM
For a podcast on the out-of-body illusion, see:
go.nature.com/9bbm4g

Ed Yong *is a freelance writer based in London who writes the blog 'Not Exactly Rocket Science'.*

1. Petkova, V. I. & Ehrsson, H. H. *PLoS ONE* **3,** e3832 (2008).
2. Guterstam, A., Petkova, V. I. & Ehrsson, H. H. *PLoS ONE* **6,** e17208 (2011).
3. van der Hoort, B., Guterstam, A. & Ehrsson, H. H. *PLoS ONE* **6,** e20195 (2011).
4. Botvinick, M. & Cohen, J. *Nature* **391,** 756 (1998).
5. Ehrsson, H. H. *Science* **317,** 1048 (2007).
6. Petkova, V. I. *et al. Curr. Biol.* **21,** 1118–1122 (2011).
7. Ionta, S. *et al. Neuron* **70,** 363–374 (2011).
8. Rosen, B. *et al. Scand. J. Plast. Recons.* http://dx.doi.org/10.1080/02844310903113107 (2009).

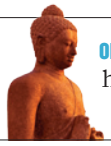
© 2011 Macmillan Publishers Limited. All rights reserved

COMMENT

SOCIAL SCIENCE Clearer ethical review guidance needed for online research **p.174**

NEUROSCIENCE What goes on in the brain when we sniff a fine red wine **p.176**

PHILOSOPHY Two books explore Buddhist views of consciousness **p.178**



OBITUARY Norman Ramsey, hydrogen masers and atomic clocks **p.182**

QUIRKY CHINA NEWS/REX FEATURES



China's unmanned spacecraft Shenzhou-8, launched in November, aims to dock with the Tiangong-1 space lab module launched in September.

Asia's space race

Greater global cooperation is needed to avert the risk of further militarization, says **James Clay Moltz**.

Asian nations are in the midst of an unspoken space race. Although China has captured most of the attention, Japan, India, South Korea and other countries in the region are also expanding their space programmes. Unlike Europe, however, where countries are cooperating within the European Space Agency (ESA), Asian nations are going it alone.

Space activity's close links to national prestige and military operations are part of the problem. Like the superpowers of the 1960s, Asian officials believe that space programmes will bring them status at home and abroad. Asian nations do not want to be seen as 'backward' or, worse, as falling behind their neighbours in science and technology. So they watch one another's space accomplishments closely, and try to match their rivals at every turn.

Tensions in Asian politics underlie today's space competition. Long-festering historical and geopolitical feuds have created hostile dyads throughout Asia: China–India, India–Pakistan, North Korea–South Korea, China–Japan and Vietnam–China, as well as lesser rivalries in Southeast Asia. Unlike Europe and the cold-war-era United States and Soviet Union, Asia has no legacy of regional security cooperation, much less arms control.

NATURE.COM
Read more about India's rise to the Moon:
go.nature.com/9ytaoz

Asia's emerging space race is spurring technological advances in the region. But competing national agendas are fostering scientific duplication, a failure to pool resources, political mistrust and increasing military tensions. These are worrisome trends that are also undercutting accomplishments made by the United States, Russia and Europe in space cooperation. The biggest fear among military analysts is that ▶

© 2011 Macmillan Publishers Limited. All rights reserved

Asia's civilian space race will turn into an arms race. Greater cooperation is needed to avert an impending confrontation.

Asia's rise in space can be tracked in the arithmetic of recent launch activity (see 'Asia's space leaders'). Last year, for the first time, the number of launches by China equalled that of the United States at 15, with only Russia ahead of them. In addition, India carried out three and Japan two. Major Asian countries are likely to increase launch activity as they attempt to deploy constellations of precision-navigation and timing satellites for both civilian and military uses.

A recent flurry of Moon missions highlights the competitive goals of Asia's space players. Japan, China and India have all conducted independent lunar-mapping programmes (Kaguya, Chang'e and Chandrayaan, respectively) since 2007. All three countries are planning follow-up missions, including landers, instrumented rovers and lunar bases. Despite this scientific redundancy, no one wants to cooperate and share the prestige from these flights.

"Asia's space relations cannot be divorced from the broader political climate in the region."

RACE TO THE TOP

Although science matters, human spaceflight garners more attention. Witness the international tumult following the flight of China's Shenzhou-5 in October 2003. With subsequent flights of two and then three taikonauts — the latter flight including a spacewalk — China has put others in Asia on notice. Its September 2011 launch of Tiangong-1, the first orbital test module for a space station planned for 2020, has raised the bar even higher.

Japan has more experience in human spaceflight, with 15 manned missions since 1992. It is a member of the International Space Station (ISS), to which it contributed the Kibo research module. Yet all of its flights have been on the US space shuttle or the Russian *Soyuz*. Japan's H-II Transfer Vehicle is now bringing cargo to the ISS. It could be adapted to carry people, but Japan has not yet committed itself to developing the necessary life-support and re-entry technology.

India has felt forced to defend its space reputation in the face of China's rise. It has recently announced that it will launch its own astronauts into orbit by 2016. In preparation, it is working with Russia to train a cadre of astronauts and to acquire the necessary technologies. But this will be an expensive and risky undertaking.

Beyond human spaceflight, all of Asia's major space-faring countries are seeking to develop supplements to the US Global Positioning System (GPS). China is again in the lead, having launched nearly one-third of its planned 35-satellite Beidou constellation. Like the GPS, Beidou will have both civilian and military applications.

Japan is in the process of building a three-satellite GPS-augmentation system called Quasi-Zenith, which will reach throughout northeast Asia. It is seeking buy-in from South Korea to manage costs.

India is planning a satellite network, called the GPS-Aided Geo Augmented Navigation System, to provide services in South Asia. Although there is some cooperation by these countries with the United States, Europe and Russia in this field, none is planned among the Asian systems.

Smaller countries are also joining the fray. Malaysia piggybacked on its purchase of Russian MiG-31 fighter planes to have its first astronaut trained by Russia and flown to the ISS in October 2007. Its rival Singapore is sending engineers and military officers to the United States for training to develop telecommunications and remote-sensing capabilities. Singapore's first domestically produced X-Sat microsatellite achieved orbit aboard an Indian booster in April.

Meanwhile, Indonesia has a well-developed satellite-communications network built with assistance from US and European companies, and in 2007 launched a remote-sensing system, built with Germany, aboard an Indian Polar Satellite Launch Vehicle rocket. Given its favourable equatorial location, Indonesia plans to develop its own launch capability by 2014.

Thailand operates several large communications satellites covering Southeast Asia. It is working with France to develop an Earth-imaging satellite and has cooperative projects with both China and Japan.

Vietnam has constructed a system of ground stations to support its US-built Vinasat-1 communications satellite and is developing its own research satellite, called Pico Dragon. Backed by Japanese development funding, Vietnam is also building a national space centre and an educational programme to support it.

Taiwan keeps tabs on China through a unique cooperative arrangement giving it access to an advanced Israeli imaging satellite as the craft passes through Asia. By 2014, Taiwan will have its own domestically built satellite, Formosat-5, for reconnaissance and Earth imaging.

The Australian government has recently established a new space organization. Its 2009 Defence White Paper announced a major push into space, including plans to develop synthetic-aperture radar capability. Australia has purchased entry into the US Wideband Global Satcom military-communications constellation, and in 2010 signed an agreement with the United States for enhanced cooperation in space tracking.

On the Korean peninsula, South Korea is using Russian Angara boosters in an effort to beat rival North Korea into space. In 2009, Yi So-yeon, a young female chemist, became the country's first astronaut, visiting the ISS aboard a Russian booster. South Korea has also developed infrastructure for satellite manufacturing and remote sensing.

Not wanting to lag behind, North Korea has twice tried to use its missile programme to place a primitive satellite into orbit, both times without success (although domestic propaganda has trumpeted otherwise).

RISING TENSIONS

Although most of Asia's space participants have focused on developing commercial space applications and military-support programmes such as reconnaissance, communications and navigation, the region's largest militaries are moving into weapons applications, causing global concern.

China's use of a ground-based interceptor to demolish one of its old satellites in January 2007 created more than 3,000 pieces of hazardous orbital debris, most of which will be in orbit until around 2050. Numerous spacecraft — including Chinese satellites — have already had to move to avoid this debris.

In response to this perceived anti-satellite threat, India has formed the Integrated Space Cell to manage its future military space assets, including satellites pledged to each of its military branches. The Singh government has said that it will develop anti-satellite weapons through a direct-ascent, missile-defence programme. This shift from its prior, exclusive focus on the civilian Indian Space Research Organisation marks a significant change in a decades-long peaceful space policy.

Until recently, all of Japan's space activities had to be 'civilian' in nature, according

ASIA'S SPACE LEADERS

Country	Space budget (yearly US$)	Civilian space personnel	Launches per year
Japan	$3.8 billion	8,300	2–3
China	$2.2 billion*	80,000*	10–15
India	$1.3 billion	32,000*	2–3
South Korea	$220 million	2,500*	0–1

*Estimated

© 2011 Macmillan Publishers Limited. All rights reserved

NASA; SIPA PRESS/REX FEATURES



Above a cloud-covered Earth, Japan's H-II Transfer Vehicle (left) successfully docked with the International Space Station in September 2009. The following year, South Korea's Naro-1 rocket (right) failed for the second time to launch, crashing just minutes after blast-off in Goheung.

to a law passed by the National Diet in 1969. But China's military space activities and the desire to monitor North Korean military developments drove the Diet, in 2008, to legislate in favour of military space activities for the first time.

Some senior Japanese officials now state that space-based weapons are conceivable for the nation, provided they are 'defensive' in nature. Japan recently announced a boost in its military space spending — for reconnaissance and early warning — despite ongoing budget woes after the earthquake and tsunami in March.

Ironically, as Asian space tensions have mounted, programmes elsewhere have grown less competitive. The United States and Russia cooperate extensively on the ISS and in commercial launch ventures. But no such peer cooperation exists among Asia's space rivals; instead, their links are skewed. Asia's leading space-faring countries engage with the United States, Europe and Russia to acquire high-level technology and training, but they prefer to work with less-developed countries to promote their own leadership, economic and security interests.

For instance, China hosts the Asia-Pacific Space Cooperation Organization (APSCO), which began operations in 2008. The group coordinates joint space-science and civil-applications research projects such as the Small Multi-mission Satellite, which is used for Earth monitoring.

China hopes that the APSCO will eventually become like ESA, but, to date, its membership comprises the less-developed countries of Bangladesh, Iran, Mongolia, Pakistan, Peru and Thailand.

Japan operates a parallel, but less-formal, organization called the Asia-Pacific Regional Space Agency Forum (APRSAF). The forum's activities include yearly conferences, training and cooperative scientific programmes such as the Sentinel Asia project for regional disaster warning and monitoring. The APRSAF has many regional participants — including a few Chinese institutes — but it is also a way to promote Japan's regional interests. There are no cooperative projects between the APSCO and the APRSAF.

MORE COLLABORATION

What can be done? Asia's space relations cannot be divorced from the broader political climate in the region. But there are solid technical, economic and environmental reasons for enhanced regional cooperation in space. What is lacking is political will.

Some data exchanges and networking have begun for disaster monitoring and tsunami warning, given recent natural catastrophes in Asia. This could be an area for enhanced cooperation between the APSCO and the APRSAF. Lunar research is another promising area. China, Japan and India are participants in the NASA-hosted Global Exploration Strategy and have expressed interest in joining the data exchanges planned under the International Lunar Network; a joint mission or hosted payload relationship could serve to reduce tensions and foster stronger linkages.

Further forward, a human spaceflight initiative by Japan or China — modelled on the US–Soviet *Apollo–Soyuz* docking in 1975 — could help to foster operational cooperation and personnel exchanges. Such contacts would promote transparency and personal connections among scientists.

Most important is the reduction of military tensions. At present, the major players lack even a basic satellite non-interference pledge like the one that was built into US–Soviet relations in 1972. A regional initiative on this score could help to reverse today's hostile trends. Similarly, the threat of orbital debris could provide the basis for talks on halting mutually damaging kinetic- and laser-weapons tests.

As an experienced and interested observer, the United States might have a part to play. Increased efforts to implement the US government's 2010 call for 'responsible behaviour' in space could be a starting point, perhaps with the United States hosting a dialogue among major space-faring nations. Although the Obama administration faces opposition in the US Congress, initiatives with China in lunar research, human spaceflight or space-traffic management could be used to draw the focus of reluctant Asian actors away from regional rivalry and towards shared, 'humankind' interests and safer norms of conduct in space.

The present course of Asian space activity risks a future collision, both literally and figuratively. Unfortunately, absent a disaster that affects all players or bold new leadership to promote cooperation, the situation seems likely to get worse before it gets better. ■

James Clay Moltz *is a professor in the Department of National Security Affairs at the Naval Postgraduate School, Monterey, CA, 93943. His latest book is* Asia's Space Race: National Motivations, Regional Rivalries, and International Risks.
e-mail: jcmoltz@nps.edu

© 2011 Macmillan Publishers Limited. All rights reserved



J. JOHANNESSEN/PANOS

Aspects of gaming cultures, such as players' conventions, are unfamiliar to many members of the ethics committees that must judge web-based privacy issues.

Open up online research

Social media hold a treasure trove of information. But the secretive methods of ethics review boards are hindering their analysis, says **Alexander Halavais**.

The mass adoption of networked communication and an emerging culture of open sharing have provided a boon for social scientists. They have opened a window through which social scientists can observe the experiences of individuals and groups that have been all-too-easily ignored by history, and present an opportunity for researchers to examine not just a few, but thousands of eyewitness accounts of an event. My own work, for example, has examined protests on Twitter, how people use the Digg website to become part of an online community, how political campaigners reach voters online and how blogs can contribute to the spread of news. The US Library of Congress, which last year acquired the archive of all public tweets made since Twitter began in 2006, announced that it will open up access to the database, although it is not yet clear when or under what rules. Through social media, social science is entering an age of 'big science'.

Collecting public musings can create ethical dilemmas for researchers. The messages may be publicly available, but the individuals writing them might not realize just how public they are, and might be surprised when their words are preserved or placed in another context. When, in 2006, sociologists from Harvard University in Cambridge, Massachusetts, collected data from Facebook for a study on students' friendships and shared interests, some people questioned whether the researchers had sufficient permission to distribute the data, and whether they had made them sufficiently anonymous. Similarly, when America Online (AOL) released a data set of some 20 million search terms from 650,000 anonymized users for the purposes of research, also in 2006, journalists and others managed to link some individuals to strings of search terms. The scandal resulted in a class-action lawsuit and the resignation of AOL's chief technology officer.

Working on large-scale public conversations is new ground for many researchers, and for most research ethics committees, also called institutional review boards (IRBs). What constitutes ethical conduct remains blurry. In the case of the Harvard Facebook study, the methodology passed ethics review, but still caused controversy. Equally troubling is the other side of the coin: many studies have been blocked by ethics review, even when they present minimal risk to the participants.

IRBs were initially created as a way to review medical research in the United States, protect against ethical misconduct and ensure that participants were made aware of the risks associated with medical experimentation. By the late 1960s, the boards were also overseeing privacy risks in social-science research, prompting the prominent US anthropologist Margaret Mead to argue to the National Institutes of Health that her field did not work with "subjects", but rather with "informants in an atmosphere of trust and mutual respect". By the mid-1990s, particularly in the United States, IRBs were overseeing groups that they used to ignore: historians, journalists and folklorists, for example. As Laurie Essig, a sociologist at Middlebury College in Vermont, put it in a blog post in August in the *Chronicle of Higher Education*: "IRBs have treated speaking with someone as equivalent to experimenting on them and have almost killed fieldwork in the process". The US model is becoming the global norm, as IRBs elsewhere begin to review studies being done in the social sciences and humanities.

The time and expense of intensive ethics review of online social science puts the brakes on such work: both slowing down

© 2011 Macmillan Publishers Limited. All rights reserved

research and restricting the sharing of data.

Steps are being taken to resolve these issues, but more needs to be done. Journals and funding agencies can, and should, help by making ethics reviews more transparent. This will demystify the ethics considerations behind such work so that researchers can learn from the successes of others. It will also encourage companies and the public to entrust their personal information to scientists. IRBs, for their part, should not apply the same standards used for studies that entail risks to health to those that may just have privacy issues, particularly when the societal perception of online privacy is in flux.

WAR STORIES

Working social scientists will agree that research conduct must be ethical, but few will extol the virtues of the IRB. War stories abound: a protocol held up because someone on the committee felt that the area of research was fruitless, or because of a spelling error, or because the research might have brought too much controversy to the campus. Even when the issues are more central to the protection of people, the standards of approval are often ambiguous and informal. High levels of scrutiny are clearly necessary for a drug trial. But scrutinizing whether gamers would be traumatized by being asked questions about dressing up as characters for conventions — to take an example from my students' research — is an issue best addressed by the researchers, who have had much more exposure to the participants and the culture being examined.

For those who research online interactions, it can be especially frustrating to have a board filled with members who have never used Facebook or played *World of Warcraft*. Although IRBs can, and sometimes do, bring in experts who can address the context of the research more directly, this happens more rarely than it should.

Members of IRBs tend to be more comfortable with some methods than others: the hypothesis-driven experiments of the psychology lab rather than the inductive work done by ethnographers, for example. The decisions of IRBs seem to be idiosyncratic and, by extension, capricious, especially when multi-site research is approved by several boards, yet held up by others. In the case of a colleague, each of two review boards insisted on having the other approve a protocol first. This can easily lead to research gridlock, and has spawned a growing industry of professional ethics-review expediters.

Calls have continually been made to improve the ethics review process. The Institutional Review Blog (www.institutionalreviewblog.com) chronicles the overzealous restrictions of IRBs in the humanities and social sciences, for example. The largest organization of ethics review professionals, Public Responsibility in Medicine and Research, provides a venue for discussing ways to improve the process. Perhaps most promisingly, in July, the federal agency that oversees IRBs in the United States — the Office for Human Research Protections — announced its intention to revise the rules for research that poses only a minimal risk to people. The proposed changes should reduce many of the burdens currently imposed on researching open discourse on the web — such as studying blog posts or tweets. It is, however, unlikely to make standards of privacy clearer to researchers, students or those who host participatory websites.

After researchers have battled their way through the IRB process, the result, an IRB approval form, is usually tucked away in a drawer. Students are often introduced to this secretive process in a cursory overview of ethics in a methods course, and sometimes by being assigned to handle the IRB process for a project. Both good and bad models of ethical research are difficult to come by for students, particularly when the work is treading less trammelled ground: ethnographic research of virtual communities, for example.

FIXING THE SYSTEM

The solution is not to do away with the IRB, but rather to make amendments that render its dysfunctions less acute.

The first step is to agree on a reasonable threshold for when IRBs should become involved — although federal oversight bodies set a minimum standard, institutions have some discretion on their specific policies. Some universities, in an overabundance of caution, require work that incurs even the most minimal risk to be reviewed by an ethics committee before it can commence, often leading to months of delays. And not all committees consider research that involves text — such as tweets or blog entries — to be in need of ethics oversight. This lack of clarity and consensus results in unnecessary oversight, taxing boards that could spend their time better on work that presents a significant risk to vulnerable people. When risks are to adults' privacy only, for example, an IRB should not need to review the protocols.

"The decisions of IRBs seem to be idiosyncratic and, by extension, capricious."

A complementary important step is to ensure that upcoming students, in all fields, have adequate ethics training. This will help to make certain that research not requiring IRB approval still has ethical thought behind it, and discourage the notion that handing a proposal over to an IRB excuses researchers from considering the ethics themselves. Ethics thinking should begin when a study is first being designed and continue through to its completion and publication, rather than constituting a single bureaucratic hurdle.

The greatest problem faced by the ethics system is secrecy. Review boards must make decisions with limited access to previous cases. Their decisions are rarely available to other IRBs or to researchers who could make productive use of their precedents — particularly in new areas of research such as online social science, in which review boards tend to have little experience and would therefore benefit most from the experience of others. One solution would be to require IRBs to be transparent in their decision-making. This, however, seems unlikely to succeed. The IRB system is conservative by design; we must look elsewhere for change.

Most government funding agencies and private foundations prominently promote open data sharing and collaboration. Were they to make the open publication of IRB protocols or ethics reflections a requirement for receiving funding, it would provide a crack in an otherwise too-secretive process. Many journals expect social-science research to have been inspected by an IRB before submission; a few even require authors to sign a statement of IRB approval. None requires that the approved protocol be provided to the journal or published. They should. Although some information would need to be redacted from such documents — including anything that might violate the privacy of participants or of the researchers — they would open a new window on ethics considerations.

Enacting these changes would not only grease the wheels of social-science research, but also help to convince the public that the work is both important to their well-being and being done in a trustworthy way. That, in turn, might make companies more willing to share their data with science. At present, researchers find it difficult to study the large data sets from Facebook or Twitter, for example, without partnering up with someone within those companies; the sites' terms of use prevent outsiders from 'scraping' large data sets.

By moving the consideration of ethical conduct beyond the localized IRB to the wider research community, we can evolve the kinds of standards and best practices that can serve to instruct not just the scholarly world, but also the wider realms of government policy and corporate practice. ■

Alexander Halavais *is president of the Association of Internet Researchers and associate professor of interactive media at Quinnipiac University, Hamden, Connecticut 06518-1908, USA.*
e-mail: alex@halavais.net

© 2011 Macmillan Publishers Limited. All rights reserved



The sense of smell is essential, coming into play in everything from personal identity to assessing the quality of foods such as this pecorino cheese.

S. SCATA/STONE/GETTY

NEUROSCIENCE

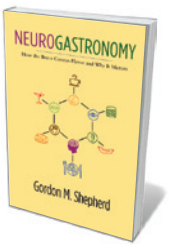
Scent and sensibility

Chris Loss savours a wide-ranging exploration of flavour that takes in evolution and physiology — and suggests it could be key to a healthier future.

The most important apparatus in the culinary arts is not cutting-edge kitchen equipment or exotic ingredients; it is the senses. As neuroscientist Gordon Shepherd describes in *Neurogastronomy*, it's through the senses that we translate the size and configuration of volatile molecules into qualitative meaning, from biochemical into flavourful.

In this succinct yet informative book, Shepherd distills more than 50 years of research — and the collective knowledge of scientists, gastronomes, chefs, anthropologists and artists — to improve our understanding of flavour. He describes the physiology and neuro-circuitry of eating in compelling detail, delving deeply into the brain's systems for flavour detection. After introducing the science, he provides snapshots of seminal academics and culinary figures past and present.

Reading *Neurogastronomy* has convinced me that the answer to the question of why we are here is 'flavour'. After all, the largest gene family in mammals codes for the primary determinants of flavour: the olfactory receptors. These genes comprise around 3% of the whole human genome — a greater percentage than that of immunoglobulin and T-cell receptors combined.



Neurogastronomy: How the Brain Creates Flavor and Why It Matters
GORDON M. SHEPHERD
Columbia University Press: 2011. 288 pp. $24.95, £16.95

Shepherd starts with the basic science of smell and taste. Smell, he explains, is a dual sense: orthonasal and retronasal. When you take a sniff of something, such as by poking your nose into a glass of wine, that is orthonasal smelling.

Retronasal smelling is more involved. When you swallow food or drink, you close your mouth and exhale through your nose. A puff of air is then pumped past the olfactory epithelium in the upper nasal cavity, where the olfactory receptors are embedded. Volatile chemicals in the air 'tickle' the receptors, which generate a signal in the olfactory bulb. This is converted into an 'odour image' that the brain recognizes as an aroma. As Shepherd explains, retronasal smell has the bigger role in flavour.

We learn that flavour is a multi-modal perception: other senses come into play too. Mouth-feel — sensing the texture of foods through receptors inside the cheek, palate and tongue — can range from springy, chunky and stiff to gritty, astringent and painful (burning, sticking or aching). Hearing and sound are also important in relation to food texture and enjoyment of flavours: studies have found that crunching raw carrots produces a pitch of 1–2 kilohertz, whereas eating crispy flatbread produces more than 5 kHz. We generally prefer "clearer and louder" food sounds, says Shepherd, who speculates that we interpret them using the brain's neocortex.

Beyond the immediate thrill of tasting chocolate, Thai curry or a red Bordeaux, flavour is an intimate part of ourselves, says Shepherd. Through it we connect to our environment, learn about our individuality and even gain insight into the origins of the human condition.

The book quotes anthropologist Richard Wrangham, who states that the advent of cooked food was key to human evolution: it required less energy to chew and increased the bioavailability of

➲ NATURE.COM
For more on the science of scent:
go.nature.com/zl6mvh

© 2011 Macmillan Publishers Limited. All rights reserved

nutrients. The energy used to physically and chemically break down food could instead be diverted towards brain development, he says. Shepherd posits that our retronasal olfactory apparatus — and the flavours detected through it — allowed for detection and identification of a broader diversity of nutrients. This process, he says, ultimately led to the establishment of cuisines, which, apart from their national and cultural connotations, are one of the key ways in which humans successfully adapt to their environments.

Understanding that flavour is largely a creation of the brain, Shepherd argues, is crucial to addressing food-related health issues facing society today, such as lowering caloric intake and reducing sodium content without compromising flavour. When the nutritional sciences are applied in the food industry, the focus is often on the effects of including or excluding certain nutrients for health. But with improvements in nutritive heft must come improvements in flavour. The emerging field of neurogastronomy recognizes that requirement, and Shepherd is just one voice in a growing chorus calling for more collaboration among flavour experts, including chefs, neuroscientists and biochemists.

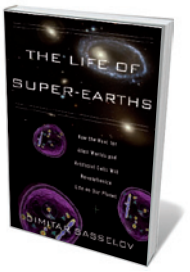
The book highlights key culinary figures past and present. Notably, the nineteenth-century French epicure and politician Jean Anthelme Brillat-Savarin, author of *The Physiology of Taste* (1825), pops up throughout to remind us that what we eat fosters critical evaluation of the natural world. Food and its flavours were the main point of inquiry for Savarin, and he pioneered some of Shepherd's major themes, basic concepts and underlying mechanisms. Along with modern gastronomic heroes such as French chef Paul Bocuse, we encounter US food-science writer Harold McGee, who has inspired and guided chefs towards more evidence-based cuisine, and has investigated the role of flavour chemistry in the science underlying deliciousness.

Shepherd ends with an adept deconstruction of the famous sensory journey of Marcel Proust in his book *Remembrance of Things Past* — specifically, the tisane of lime-flower tea and the madeleine cake that sets Proust off on his epic memory-fest and reveals the hard-wired connectivity of flavour stimuli to memory, emotion and critical thought.

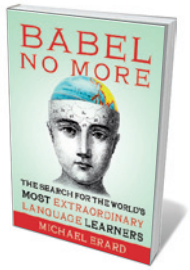
If flavour is the trigger for an act of the imagination as extraordinary as Proust's magnum opus, what else might it do? Shepherd makes an excellent case for neurogastronomy as an important cross-disciplinary field that is likely to motivate a variety of imperatives for our health and well-being. ■

Chris Loss *is a food scientist and director of menu research and development at The Culinary Institute of America, Hyde Park, New York 12538, USA.*
e-mail: c_loss@culinary.edu

Books in brief



The Life of Super-Earths: How the Hunt for Alien Worlds and Artificial Cells Will Revolutionize Life on Our Planet
Dimitar Sasselov BASIC BOOKS *240 pp. $24.95 (2012)*
Planet hunters have been busy bagging their quarries since 1995: known extrasolar planets now number around 600. In this short, sharp look at the subset called 'super-Earths' — rocky or oceanic, but more massive than Earth — astronomer Dimitar Sasselov gives us the science and the speculation about life on other worlds. He suggests that the Copernican revolution, which demoted Earth from its position at the centre of everything, could be brought full circle by new findings from synthetic biology and planetary science.



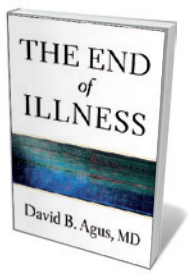
Babel No More: The Search for the World's Most Extraordinary Language Learners
Michael Erard FREE PRESS *320 pp. $25.99 (2012)*
Polyglottism has always amazed. But are hyperpolyglots — who speak many languages — neurological oddities, swots or genetically predisposed? Writer Michael Erard's tale of tongues includes such figures as Bolognese priest Giuseppe Mezzofanti, alleged master of 40 languages, and British explorer Richard Burton, who spoke 29 (and 11 dialects). Erard examines sections of multilinguists' brains and tracks down today's hyperpolyglots. He concludes that they are a "neural tribe" — ambitious and willing to reshape their brains.



Golden Holocaust: Origins of the Cigarette Catastrophe and the Case for Abolition
Robert N. Proctor UNIVERSITY OF CALIFORNIA PRESS *779 pp. $49.95 (2012)*
A century of hype has made tobacco the most popular drug on Earth. Although smokers may feel they are working the glamour of film noir, the facts about this lethal habit just get grimmer.
For his monumental and sobering indictment, science historian Robert Proctor dug through piles of recently released industry documentation to uncover the activities that lured many scientists into its mill of denial. A tale of giant profits, decades of secrecy over the links with cancer, useless filters and more.



The Sacred Headwaters: The Fight to Save the Stikine, Skeena, and Nass
Wade Davis and Carr Clifton GREYSTONE BOOKS *160 pp. $45 (2012)*
Splayed next to southern Alaska, Canada's Sacred Headwaters region is a vast panorama of mountains, salmon rivers and canyons criss-crossed with the trails of caribou, grizzlies and mountain goats. Thousands of First Nations people live there. But as anthropologist and ethnobotanist Wade Davis explains, it could become a war zone. Corporations are queuing up to develop the region, halted only by First Nations activists and a government moratorium that lasts until 2013. Carr Clifton's haunting photographs evoke what's at stake.



The End of Illness
David B. Agus SIMON & SCHUSTER *352 pp. £14.99 (2012)*
The billions of dollars spent on medical research have failed to vanquish cancer and other serious diseases. Solutions, argues medical oncologist David Agus, depend on a systemic, whole-body approach rather than a reductionist one. Emphasizing prevention and tailored treatment, Agus advocates the combination of techniques such as proteomics with pragmatic routines based on robust research. Avoiding risky habits and introducing lifestyle changes such as taking aspirin are small steps that could pay dividends, he says.

© 2011 Macmillan Publishers Limited. All rights reserved

nutrients. The energy used to physically and chemically break down food could instead be diverted towards brain development, he says. Shepherd posits that our retronasal olfactory apparatus — and the flavours detected through it — allowed for detection and identification of a broader diversity of nutrients. This process, he says, ultimately led to the establishment of cuisines, which, apart from their national and cultural connotations, are one of the key ways in which humans successfully adapt to their environments.

Understanding that flavour is largely a creation of the brain, Shepherd argues, is crucial to addressing food-related health issues facing society today, such as lowering caloric intake and reducing sodium content without compromising flavour. When the nutritional sciences are applied in the food industry, the focus is often on the effects of including or excluding certain nutrients for health. But with improvements in nutritive heft must come improvements in flavour. The emerging field of neurogastronomy recognizes that requirement, and Shepherd is just one voice in a growing chorus calling for more collaboration among flavour experts, including chefs, neuroscientists and biochemists.

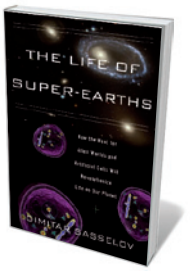
The book highlights key culinary figures past and present. Notably, the nineteenth-century French epicure and politician Jean Anthelme Brillat-Savarin, author of *The Physiology of Taste* (1825), pops up throughout to remind us that what we eat fosters critical evaluation of the natural world. Food and its flavours were the main point of inquiry for Savarin, and he pioneered some of Shepherd's major themes, basic concepts and underlying mechanisms. Along with modern gastronomic heroes such as French chef Paul Bocuse, we encounter US food-science writer Harold McGee, who has inspired and guided chefs towards more evidence-based cuisine, and has investigated the role of flavour chemistry in the science underlying deliciousness.

Shepherd ends with an adept deconstruction of the famous sensory journey of Marcel Proust in his book *Remembrance of Things Past* — specifically, the tisane of lime-flower tea and the madeleine cake that sets Proust off on his epic memory-fest and reveals the hard-wired connectivity of flavour stimuli to memory, emotion and critical thought.

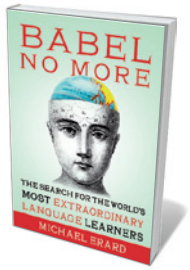
If flavour is the trigger for an act of the imagination as extraordinary as Proust's magnum opus, what else might it do? Shepherd makes an excellent case for neurogastronomy as an important cross-disciplinary field that is likely to motivate a variety of imperatives for our health and well-being. ■

Chris Loss *is a food scientist and director of menu research and development at The Culinary Institute of America, Hyde Park, New York 12538, USA.*
e-mail: c_loss@culinary.edu

Books in brief



The Life of Super-Earths: How the Hunt for Alien Worlds and Artificial Cells Will Revolutionize Life on Our Planet
Dimitar Sasselov BASIC BOOKS *240 pp. $24.95 (2012)*
Planet hunters have been busy bagging their quarries since 1995: known extrasolar planets now number around 600. In this short, sharp look at the subset called 'super-Earths' — rocky or oceanic, but more massive than Earth — astronomer Dimitar Sasselov gives us the science and the speculation about life on other worlds. He suggests that the Copernican revolution, which demoted Earth from its position at the centre of everything, could be brought full circle by new findings from synthetic biology and planetary science.



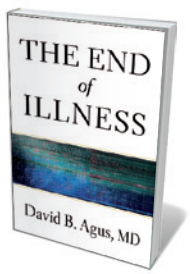
Babel No More: The Search for the World's Most Extraordinary Language Learners
Michael Erard FREE PRESS *320 pp. $25.99 (2012)*
Polyglottism has always amazed. But are hyperpolyglots — who speak many languages — neurological oddities, swots or genetically predisposed? Writer Michael Erard's tale of tongues includes such figures as Bolognese priest Giuseppe Mezzofanti, alleged master of 40 languages, and British explorer Richard Burton, who spoke 29 (and 11 dialects). Erard examines sections of multilinguists' brains and tracks down today's hyperpolyglots. He concludes that they are a "neural tribe" — ambitious and willing to reshape their brains.



Golden Holocaust: Origins of the Cigarette Catastrophe and the Case for Abolition
Robert N. Proctor UNIVERSITY OF CALIFORNIA PRESS *779 pp. $49.95 (2012)*
A century of hype has made tobacco the most popular drug on Earth. Although smokers may feel they are working the glamour of film noir, the facts about this lethal habit just get grimmer.
For his monumental and sobering indictment, science historian Robert Proctor dug through piles of recently released industry documentation to uncover the activities that lured many scientists into its mill of denial. A tale of giant profits, decades of secrecy over the links with cancer, useless filters and more.



The Sacred Headwaters: The Fight to Save the Stikine, Skeena, and Nass
Wade Davis and Carr Clifton GREYSTONE BOOKS *160 pp. $45 (2012)*
Splayed next to southern Alaska, Canada's Sacred Headwaters region is a vast panorama of mountains, salmon rivers and canyons criss-crossed with the trails of caribou, grizzlies and mountain goats. Thousands of First Nations people live there. But as anthropologist and ethnobotanist Wade Davis explains, it could become a war zone. Corporations are queuing up to develop the region, halted only by First Nations activists and a government moratorium that lasts until 2013. Carr Clifton's haunting photographs evoke what's at stake.



The End of Illness
David B. Agus SIMON & SCHUSTER *352 pp. £14.99 (2012)*
The billions of dollars spent on medical research have failed to vanquish cancer and other serious diseases. Solutions, argues medical oncologist David Agus, depend on a systemic, whole-body approach rather than a reductionist one. Emphasizing prevention and tailored treatment, Agus advocates the combination of techniques such as proteomics with pragmatic routines based on robust research. Avoiding risky habits and introducing lifestyle changes such as taking aspirin are small steps that could pay dividends, he says.

© 2011 Macmillan Publishers Limited. All rights reserved

PHILOSOPHY

Meditation in mind

Daniel Stoljar contemplates two contrasting takes on science and Buddhism.

Buddhism is a distinctive world religion. It lacks an 'omnigod' — an omnipresent, omniscient, all-powerful creator — and a notion of humans as complexes of physical bodies and souls that ascend to heaven after death. Could it be mingled with the scientific culture of the twenty-first century to produce a new philosophical outlook on the world, the mind and our values? That idea lies at the heart of these two contrasting books on Buddhism and science.

In *The Bodhisattva's Brain*, philosopher of mind Owen Flanagan wants to change Buddhism to fit better with the scientific world view. In *Meditations of a Buddhist Skeptic*, Buddhist scholar and monk B. Alan Wallace wants to alter the scientific world view so that it meshes better with Buddhism. Both tease out the contrasts between Buddhism and materialism — and both stumble when it comes to defining the latter.

For Flanagan, the world is fundamentally physical, and thus explicable by natural science, at least in principle. But Buddhism seems inconsistent with this materialistic world view. It may have no place for a creator god or ascending souls, but as Flanagan notes, it is "opulently polytheistic insofar as spirits, protector deities, ghosts, and evil spirits abound". He points out, too, that in east and southeast Asia, a belief in rebirth among Buddhists is as common as a belief in heaven among North Americans. So Flanagan sets out to 'naturalize' Buddhism: to see what Buddhism would look like without the "hocus pocus" (as he cheerfully puts it).

The Bodhisattva's Brain: Buddhism Naturalized
OWEN FLANAGAN
MIT Press: 2011. 272 pp. $27.95, £19.95

Meditations of a Buddhist Skeptic: A Manifesto for the Mind Sciences and Contemplative Practice
B. ALAN WALLACE
Columbia University Press: 2011. 304 pp. $27.95, £18.95

The result is a wide-ranging discussion of the neural and cognitive basis of mental states such as meditation and the achievement of enlightenment or nirvana, which are central to Buddhism. Flanagan outlines a plausible moral philosophy based on an idea that he takes from Aristotle but reinterprets in the light of Buddhist teaching: eudaemonia, a sort of happiness that, in Flanagan's view, is the proper aim of a good life.

Wallace adopts a contrary view, even urging the many scientists and philosophers who embrace materialism to change their minds. His book is in part a compelling and clear statement of key Buddhist ideas, but its main point is to advocate a distinction between science and a materialist interpretation of it. Materialist science, Wallace thinks, cannot get to grips with the reality of consciousness, free will or values. Science inspired by Buddhist experience might.

In both books, however, the concept of materialism remains blurry. Wallace notes what materialism isn't: the ancient 'atoms and the void' notion of the Greek philosopher Democritus, which is inconsistent with modern physics. But rather than clarifying what it is, he uses materialism as a placeholder for his dislikes — variously, a nihilistic rejection of moral values, a desire for more possessions and the denial of consciousness and the mind altogether. A more explicit statement from Wallace about what he is rejecting would have given his manifesto more muscle.

Flanagan's discussion, in my view, has a similar flaw. In contemporary philosophy of the mind and science, materialism is sometimes seen as the view that neuroscience, particularly at the cellular level, can explain consciousness. But Flanagan instead advocates what he calls "the natural method" for this task. Roughly, this research approach involves gathering subjective data, such as first-person narratives, and then using ideas or theories from cognitive and neural science to explain what is going on in the brains or minds of the conscious person. If that is materialism, it is hard to see anybody disagreeing with it, Buddhist or not.

All this said, it is true that science has yet to produce good explanations of consciousness, value and free will. The suggestion brought to the fore by Flanagan and Wallace — that Buddhism may be a source of insight in these areas — is a welcome and tantalizing one. ■

Daniel Stoljar *is professor of philosophy at the Australian National University, Canberra, ACT 0200, Australia, and author of* Physicalism.
e-mail: daniel.stoljar@anu.edu.au

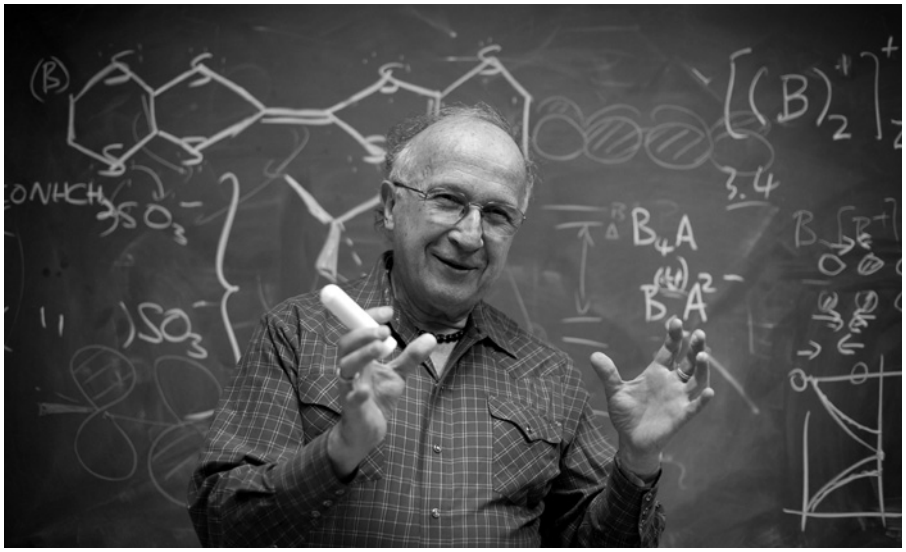


W. MARR/PANORAMIC IMAGES/ROBERT HARDING PICTURE LIBRARY

Buddhism, which reached Borobudur in Java, Indonesia, by at least the ninth century AD, could provide insight into consciousness and free will.

© 2011 Macmillan Publishers Limited. All rights reserved

M. ZENKEWICZ



Q&A Roald Hoffmann

Chemical connector

Theoretical chemist, poet and playwright Roald Hoffmann won a Nobel prize in 1981 for his work on how molecules change as they react. As the International Year of Chemistry comes to a close and he releases two books, Hoffmann talks about language, ethics and the sublime.

Why edit a volume of essays about the sublime?
Beyond the Finite is a collection by various authors writing on the role of the sublime in art and science, co-edited with Iain Boyd White. Most scientists think that the sublime is just emotional excess or, God forbid, the religious creeping into what we do. This is silly. Scientists are confronted with astounding novelty. Why not embrace the wonder one feels when looking at the ribosome, or at the beauty of a molecule shaped like a Ferris wheel or at a distant dust cloud across a galaxy? Why not accept that point at which one feels one really knows something that is deep and universal? Why not be at peace with the sublime?

What topics do you cover in the other book?
Roald Hoffmann on the Philosophy, Art and Science of Chemistry, edited by Jeffrey Kovac and Michael Weisberg, collects my writings on a number of subjects. Among these are what goes on under the surface of an article in a chemistry journal, how the philosophy of science would be different if it was done by chemists, and strategies for improving teaching. I talk of the special relationship between handicrafts and chemistry. And I fight reductionism all the way.

When did you become a writer?
I leapt into poetry from science at 40, then filled in the ground in between. I started writing essays and philosophical articles at around 50, and plays after 60. My years of getting scientific articles rejected were good training for submitting poems. It took me seven years to get a poem accepted, and I still have difficulty getting them published.

Beyond the Finite: The Sublime in Art and Science
EDITED BY ROALD HOFFMANN AND IAIN BOYD WHITE
Oxford University Press: 2011. 208 pp. $24.95, £15.99

Roald Hoffmann on the Philosophy, Art, and Science of Chemistry
EDITED BY JEFFREY KOVAC AND MICHAEL WEISBERG
Oxford University Press: 2012. 448 pp. $35, £22.50

What do you research?
Early in my career, I found a method to explain the ways in which electrons move in molecules. As a theorist, I have moved through chemistry and am now focusing on solid-state physics and conductors — and building bridges between these subfields. I have explained the shape, colour, spectra, electrical properties and reactivity of molecules using simple ideas about how electrons move in their orbitals, and I have tried to predict molecules that haven't yet been made. Chemists will tell you that I am good at devising verbal explanations and extracting understanding out of poor calculations.

You call your work "applied theoretical chemistry". What does that mean?
I'm a theorist, but I take my inspiration from experiment. I don't build castles in the sky. I don't believe in working on 'important' problems. I would rather solve the many little puzzles that I see in the literature — strange reactivities, or why one molecule is bent when another is straight. Underneath is my philosophy that if I look at hundreds of smaller problems in chemistry, and keep in mind the connections that must be there, I will see the chemical universe.

Should chemists take the blame for how the molecules they have created are being used?
We should not take the blame. But as we create the new, we should ask whether our creation — be it a molecule, a gun or a poem — can hurt people. Without that question, the act of creation is incomplete. My play *Should've* tries to deal with this question. What responsibility do we have when, despite our best intentions, something we made that appears harmless, or even good, is abused by others? Not much, but some. Here we enter the realm of Greek tragedy. This is not to set scientists on a pedestal, but to make them see that they are no more absolved of ethical choices than other people.

Your play *Oxygen* was staged across Europe this autumn. What does it portray?
Written with chemist and author Carl Djerassi, *Oxygen* comes out of the story of one week in 1774 when Antoine Lavoisier was on the verge of understanding modern chemistry. All he needed was the discovery of oxygen. Suddenly he is faced with Joseph Priestley and Carl Wilhelm Scheele, who have each discovered the element but don't understand what they have discovered. How does Lavoisier react to that? He ignores Scheele and tries to take credit away from Priestley. Then there is Madame Lavoisier, who deserves her own opera. So, who discovered oxygen? We have these romantic notions of the past, but there were disputes over credit then, just as there are now. One has to admire the Nobel committee for having the moral strength to make its decisions.

What keeps you interested in chemistry?
Chemistry is closer to the human scale than physics. Neither subatomic particles nor the shape of the Universe will affect you much. But if you are prescribed a drug, you want it to work. Molecules are like humans in some ways: they can hurt or heal us. They are constantly moving, both natural and unnatural, wondrous in their complexity. A journalist once asked me when we will discover the next element. I told him that it is not about new atoms, but rather the molecules that a chemist can build from them. As with toy blocks, it is what the child's imagination does with them that counts. ■

INTERVIEW BY JASCHA HOFFMAN

© 2011 Macmillan Publishers Limited. All rights reserved

Correspondence

Violence: the role of society and state

I question Steven Pinker's optimistic prediction of a continuing decline in interpersonal violence, which is based on supposed correlations of mortality data with improved reasoning ability and IQ (*Nature* **478,** 309–311; 2011).

Even if we disregard possible biases in the mortality data and the problems of quantifying reason and IQ over long time spans or across cultures, a sudden shift in innate individual attributes since the mid-twentieth century is unlikely. Pinker's contention that "modern societies have been getting smarter" makes a further dubious leap from an individual to a social intelligence.

More-concrete social features should be considered. The societies in which the decline of interpersonal violence is most marked have become more organized and affluent, with less contention over resources, and violence has mostly been deferred to the state. We know that this deferral can lead to violent events that are more erratic and severe — and even ideological and irrational.

Primitive terrorism and revenge can be blamed for tens of thousands of casualties in Iraq and Afghanistan. We have had the nuclear technology of violence since the Second World War. Occasions on which nuclear weapons were actually used (with some 200,000 civilian casualties), or were close to being used (in the Korean and Vietnam wars and in Cuba), indicate that the statistics for mortality by violence are unstable owing to the escalation and sensitivity of trigger points.

Rather than an irrational optimism, we should adopt a rational, risk-assessment approach that better recognizes the instability of organized and powerful social violence.

Gary G. Nelson *Troy, New York, USA. garygnelson@gmail.com*

Violence: intuition is important too

Steven Pinker argues that humans have been short of rationality throughout most of history, and suggests we may now be putting that right (*Nature* **478,** 309–311; 2011). But intuition — the gut feelings that range from raw instinct to common sense — is important too. Reasoning and intuition together lead to wisdom, but either without the other is dangerous.

Intuition derives from various subconscious responses, presumably largely evolved, that lie at the root of our moral and aesthetic sense. A healthy mind conducts a perpetual dialogue between reasoning and intuition.

Rationality without moral and aesthetic constraint has led us to nuclear war and eugenics and, more recently, to the crude logic of neoliberalism, industrialized agriculture and overconsumption.

This, of course, is where religion comes in. Although it sometimes loses sight of its own *raison d'être*, its core function is to refine our intuitions.

Colin Tudge *Oxford, UK. colin@colintudge.co.uk*

Europe must address research misconduct

In the United States, the Office of Research Integrity and the National Science Foundation both impose legal sanctions for scientific fraud in publicly funded research. The European Commission has no equivalent measures in place. Strict regulations and policy procedures to combat misconduct (see, for example, go.nature.com/a1znzb) are needed before Europe's ambitious Horizon 2020 research-funding programme is introduced in 2014 (*Nature* **478,** 16; 2011).

At present, the programme contains no provision for scientific misconduct. Even the guidelines of the European Code of Conduct for Research Integrity make no detailed recommendations for dealing with fraudulent research (see go.nature.com/juubnq).

Under the Horizon 2020 proposals, grant applicants should sign an ethics declaration confirming that they abide by the European Code of Conduct and will cooperate in and support investigation into suspected violations and alleged misconduct.

Creation of a European agency for research integrity would guarantee that any work funded by Horizon 2020 meets explicit standards. For example, misconduct investigations could be carried out by agency ombudsmen and by ad hoc research-integrity committees, based on the ethics review panels of the current Seventh Framework Programme for research. The agency would listen to whistleblowers and take over investigations impeded by institutional reluctance or by unmanageable conflicts of interest. It could also bar institutions from receiving future grants from the European Commission if they failed to cooperate.

Xavier Bosch *Department of Internal Medicine, Hospital Clinic, University of Barcelona, Spain. xavbosch@clinic.ub.es*

Caution over Iran's nuclear plans

Your assertion that the most recent International Atomic Energy Agency's report on Iran's nuclear programme uses "old intelligence" to provide "the sharpest picture yet of the weapon that Iran hopes to develop" (*Nature* **479,** 282; 2011) warrants some contextual explanation.

Some analysts say that this intelligence dates from the turn of the millennium. This is because the latest US National Intelligence Estimates of Iran's nuclear programme (one issued in 2007 and another earlier this year), which represent a consensus of all 16 US intelligence agencies, concluded that Iran ended its efforts to build a nuclear weapon in 2003 (see go.nature.com/ms1sq7).

Many of the same political elements that erroneously claimed that Iraq had nuclear weapons are the ones now insisting that Iran is trying to build a nuclear bomb. This time around, let's be sure of the facts.

John W. Grula *Carnegie Observatories, Pasadena, California, USA. jgrula@obs.carnegiescience.edu*

Canny liars score by tricking themselves

Robert Trivers makes another significant contribution to evolutionary and behavioural biology in his latest book about deceit and self-deception (*Nature* **478,** 314–315; 2011).

The idea that self-deception is beneficial because it makes it easier to deceive others has also been noted by other observers of human nature. Writer Mark Twain declared: "When a person cannot deceive himself, the chances are against his being able to deceive other people," and author Elizabeth Bear more recently advised: "The secret to getting away with lying is believing with all your heart."

Mark A. Davis *Macalester College, Saint Paul, Minnesota, USA. davis@macalester.edu*

CORRECTION

Late editorial changes to A. Lescroël and D. Grémillet's Correspondence (*Nature* **479,** 299; 2011) imply that the Ross Sea is already under absolute protection against fishing, which it is not.

© 2011 Macmillan Publishers Limited. All rights reserved

Norman Ramsey

(1915–2011)

Physicist behind precision spectroscopy and atomic clocks.

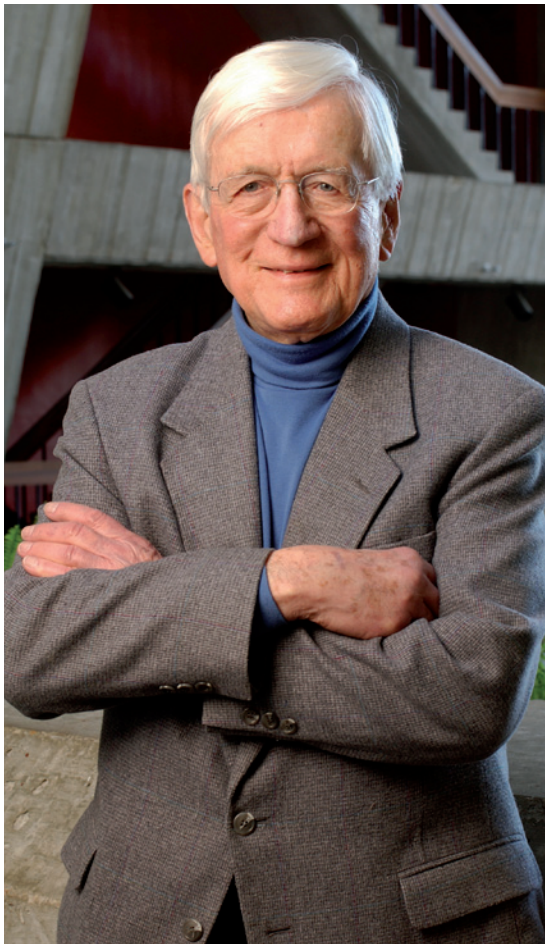
In the fields of atomic and molecular physics, Norman Ramsey's name is synonymous with a technique for precision spectroscopy, which earned him a share of the 1989 Nobel prize. Beyond the lab, he was also respected as a scientific statesman, admired for his fairness and integrity, his energy and unfailing good humour. Over his long career he inspired generations of physicists, including the 84 graduate students whom he supervised.

Ramsey, who died aged 96 on 4 November, was born in 1915 in Washington DC, the son of an Army officer and a mathematics teacher. He entered Columbia University, New York, in 1931, majoring in engineering and then shifting to mathematics. On graduation, he studied physics at the University of Cambridge, UK, working for two years alongside pioneers of nuclear physics such as J. J. Thomson, Ernest Rutherford and James Chadwick.

Returning to Columbia in 1937, Ramsey became a graduate student in the group of Isidor Rabi. Just a few months later, Rabi invented a revolutionary method for studying the structure of atoms and molecules, using streams of irradiated molecules in a vacuum. Ramsey's research blossomed: he shared in the discovery that the deuterium nucleus (a proton and neutron) is not spherical, implying that nuclear forces do not act along the line between the nuclei.

Shortly after receiving his PhD in 1940, Ramsey joined the war effort at the MIT Radiation Laboratory in Cambridge, Massachusetts, developing radar systems; later he worked on the Manhattan Project. Following the war, Ramsey returned to Columbia. With Rabi, he helped found Brookhaven National Laboratory in Upton, New York, serving as the first chair of its physics department. In 1947, Ramsey moved to Harvard University, where he remained for the rest of his career as the Higgins Professor of Physics.

At Harvard, Ramsey undertook a series of classic studies on nuclear and magnetic interactions within molecules, which helped to lay the foundations of the theory of 'chemical shifts' for nuclear magnetic resonance and magnetic resonance imaging, as well as revealing many properties of nuclei, such as their magnetism. He is best known for inventing the separated oscillatory field method — today known simply as the Ramsey method — in which particles are subjected to two pulses of temporally separated radiation, improving the resolution with which their energy structure can be measured. The method is now ubiquitous in high-precision and quantum physics, and is an essential ingredient in many atomic clocks.



R. HAHN/FERMILAB

In another breakthrough, Ramsey and his student and colleague, Daniel Kleppner, found a way to confine hydrogen atoms for a relatively long time, typically a second, thereby enabling the atoms to radiate into a resonant microwave cavity. The result was the hydrogen maser, which emits stable radiation at the atoms' resonant frequency (a wavelength of 21 centimetres) and today has a key role in the Global Positioning System and in radio astronomy. The maser was also used in 1980 to demonstrate precisely Albert Einstein's prediction that clocks run at different rates in different gravitational potentials.

As a scientific statesman, Ramsey served as first assistant secretary general for science in NATO in 1958 and 1959, establishing summer schools and fellowships that helped to educate a generation of post-war European physicists. He founded the Harvard cyclotron, and later chaired the Harvard–MIT Cambridge Electron Accelerator. For 16 years he served as president of the University Research Association, which managed the construction and operation of Fermilab in Batavia, Illinois.

During Senator Joseph McCarthy's anti-communist witch-hunt of the early 1950s, Ramsey appeared on television to rebut attacks on his colleague Wendell Furry and on Harvard University. Although he disagreed with Furry's political views, he said that "true freedom gives a man not only the right to make a right choice but also the freedom to make a wrong choice". In the 1980s, he chaired the National Research Council committee that concluded that acoustic evidence did not support the idea that a second gunman was involved in the assassination of President John F. Kennedy.

When I was one of Ramsey's graduate students, he was heavily occupied with Fermilab. Yet he seldom, if ever, missed the weekly group meetings with students at Harvard, always offering his encouragement. Even in his later years, when he wound down his research group, he continued to lecture and participate in conferences, attending almost every session and engaging with students and senior scientists alike.

Ramsey's enthusiasm and energy were legendary. At a group outing to a beach on a cold and windy day during my graduate-school days, he was the only one in the water — trying out his surfboard. And in his eighties he trekked through Nepal with his wife, hiking for miles with a severed Achilles tendon.

Norman Ramsey was a towering figure in physics. Through his accomplishments, community service, warmth and high ethical standards, he was an inspiration to many. ■

David Wineland *is a physicist in the Time and Frequency Division of the National Institute of Standards and Technology (NIST) in Boulder, Colorado 80305, USA.*
e-mail: david.wineland@nist.gov

© 2011 Macmillan Publishers Limited. All rights reserved

ORGANIC CHEMISTRY

A radical approach to diversity

Compounds containing the trifluoromethyl group have many uses, but their isomers must often be made using different multi-step routes. Two studies now show how several isomers can be made by the same route. SEE LETTER P.224

ANDREW T. PARSONS &
STEPHEN L. BUCHWALD

The incorporation of fluorine atoms into organic molecules has had a profound impact on the development of new products for numerous industries. This is largely due to the useful properties that fluorine imparts to organic molecules, such as stability and lipophilicity, and the fact that it can modify molecular polarity. Among its many applications, fluorine incorporation has been used in pharmaceutical development to prevent molecules from being metabolized too quickly, thereby allowing a drug to act before it is cleared from the body[1]. Fluorine has also played an integral part in the development of agrochemicals and high-performance materials[2]. Two papers, one by Ji *et al.*[3] in *Proceedings of the National Academy of Sciences* and the other by Nagib and MacMillan[4] on page 224 of this issue, now report innovative methods for introducing an important fluorinated group — the trifluoromethyl group, CF_3 — into molecules.

Appending fluorinated groups to small organic molecules has long posed a considerable challenge to synthetic organic chemists. The trifluoromethyl group has been particularly difficult to install, in part because the reactive intermediates that are generated during trifluoromethylation reactions are unstable under the conditions necessary for the reactions to proceed. The harsh protocols typically required for these reactions can limit the substrates that can be used and/or cause side-product formation.

Reports highlighting efforts to use milder methods that avoid the common problems associated with trifluoromethylation chemistry have therefore recently garnered great attention[5]. Although much progress has been made, many methods continue to rely on the use of 'pre-functionalized' reactants containing chemical groups that, in conjunction with transition-metal catalysts or a promoter reagent, direct trifluoromethylations to particular sites in the reactant (Fig. 1a). Alternatively, structurally complex trifluoromethyl-containing molecules can be made from simpler starting materials that already contain a trifluoromethyl group (Fig. 1b). But using these methods comes at a price: the preparation of pre-functionalized and trifluoromethyl-containing starting materials is time-consuming and costly in terms of chemicals used and waste disposal, and overall is inefficient. Furthermore, when access to a diverse range of trifluoromethylated molecules is desired, each molecule requires its own pre-functionalized starting material to be prepared, amplifying the inefficiencies so that they become increasingly prohibitive.

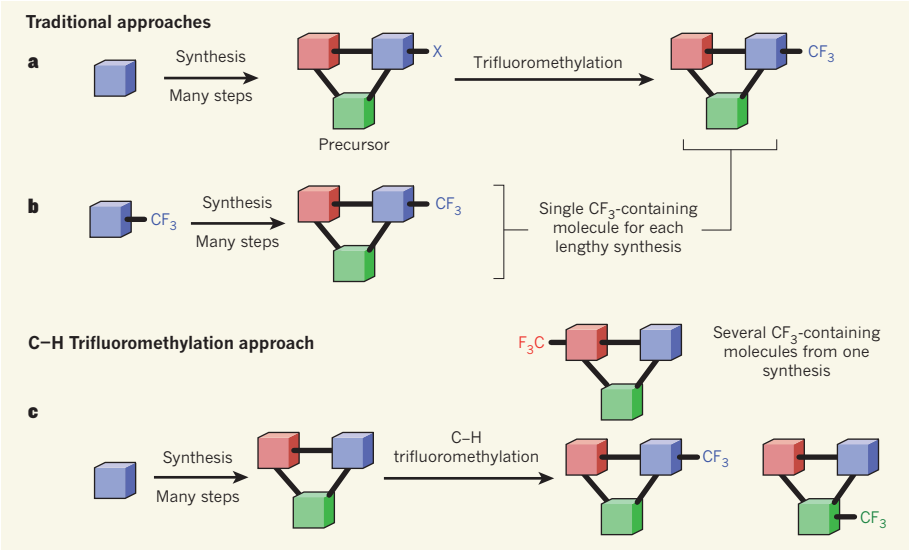


Figure 1 | Approaches for making trifluoromethylated compounds. **a**, One traditional approach for making compounds appended with trifluoromethyl (CF_3) groups involves making a precursor of each isomer that contains a surrogate group (X), which is then converted into a trifluoromethyl group at the end of the synthesis. A different multi-step synthesis is required to make each isomer; only one such route is shown. **b**, The second traditional approach starts from compounds that already contain a trifluoromethyl group. Again, a different multi-step synthesis is required to make each isomer, only one of which is shown. **c**, Two papers[3,4] now report a different approach: a structurally complex molecule is made, which is then converted into several isomers of the trifluoromethylated analogue in a single step. This strategy reduces the time and labour required to make several trifluoromethylated analogues of structurally complex molecules, and minimizes the overall waste produced.

Despite their drawbacks, existing chemoselective (site-selective) methods are often useful when the best trifluoromethylated molecule for a given application is known. But what happens when the optimal site for trifluoromethylation in a molecule is not known, as is often the case for those working in fields such as drug discovery and agrochemical research? In many instances, it is desirable to append trifluoromethyl groups to a molecule at as many positions as possible, to make lots of different isomers. This approach maximizes the diversity of the molecules obtained and increases the probability of discovering a molecule that has desirable chemical properties and biological activity.

The reports by Ji *et al.*[3] and Nagib and MacMillan[4] highlight their efforts to develop versatile methods for the rapid preparation of a diverse range of trifluoromethyl-containing molecules. Both groups sought a general protocol that provides access to trifluoromethylated molecules without the need for pre-functionalized substrates. They targeted a chemical transformation, termed C–H trifluoromethylation, in which the hydrogen of a carbon–hydrogen (C–H) bond is replaced with a trifluoromethyl group. Although ambitious, this approach would provide a large payoff in efficiency by allowing trifluoromethyl groups to be installed at late stages of

© 2011 Macmillan Publishers Limited. All rights reserved

synthetic routes, without first preparing pre-functionalized or trifluoromethyl-containing substrates. It would therefore decrease the number of synthetic operations necessary to access a variety of trifluoromethylated molecules (Fig. 1c).

The authors[3,4] recognized that the development of a successful method hinged on identifying a practical trifluoromethyl source. Both groups independently surmised that the reliable generation of highly reactive trifluoromethyl radicals ($^{\bullet}CF_3$) as intermediates might allow the direct trifluoromethylation of non-functionalized substrates, in what Ji *et al.*[3] call an "innate trifluoromethylation". In other words, the inherent reactivity of the substrates would dictate the site (or sites) at which the trifluoromethylation reaction occurs.

The two groups pursued different approaches to develop their own mild, efficient methods to access reactive trifluoromethyl-radical intermediates. Ji *et al.* developed a transition-metal-free approach that relies solely on air- and moisture-stable reagents (which are easier to handle than commonly used organometallic reagents and transition-metal catalysts that react with air and moisture). They determined that the Langlois reagent[6] ($NaSO_2CF_3$) serves as a reliable source of $^{\bullet}CF_3$ when treated with *t*-butylhydroperoxide (a commonly used reagent in radical chemistry). Nagib and MacMillan[4], however, chose triflyl chloride ($ClSO_2CF_3$) as their source of $^{\bullet}CF_3$ (refs 7, 8). They generated the radicals simply by shining light from a household fluorescent bulb onto a solution of triflyl chloride in the presence of a light-activated ruthenium catalyst. Both methods[3,4] generated $^{\bullet}CF_3$ in the presence of a diverse range of substrates, and provided access to an array of structurally complex trifluoromethylated molecules. Furthermore, both methods worked reliably even for substrates bearing potentially sensitive functional groups (those whose presence in a molecule often adversely affects the course of a range of chemical reactions, including traditional trifluoromethylation reactions).

To demonstrate the versatility of their methods, both groups conducted trifluoromethylations of common pharmaceuticals and natural products. Ji *et al.*[3] successfully achieved C–H trifluoromethylations of the natural-product derivative dihydroquinine and the billion-dollar drug Chantix (varenicline, used to combat nicotine addiction), providing products bearing a single trifluoromethyl group. These examples are significant because accessing CF_3–dihydroquinine and CF_3–Chantix using previously available trifluoromethylation methods would require laborious, multi-step total syntheses. Notably, the reactions were chemoselective, even though the starting materials contained multiple reactive sites.

Similarly, Nagib and MacMillan[4] subjected the cholesterol-lowering drug Lipitor (atorvastatin) to their reaction conditions. This experiment resulted in the non-selective formation of three chromatographically separable isomers of CF_3–Lipitor, in which the trifluoromethyl group was appended to different carbons in the molecule. The improved efficiency of this approach for accessing new chemical entities cannot be overstated, given that preparation of these isomers using traditional methods would require three parallel total syntheses, which would be extremely time- and labour-intensive.

Although the new methods[3,4] constitute a major contribution to the field of trifluoromethylation chemistry, many challenges remain. For example, it would be desirable to have precise control over the site at which C–H trifluoromethylation occurs. This might be accomplished through the development of reaction conditions that combine the best attributes of the currently available methods. Complementary protocols that selectively produce different trifluoromethyl-containing molecules from a single substrate would also be advantageous, because it would avoid the often difficult task of separating mixtures of isomers generated as products. Ji *et al.*[3] report their initial efforts towards addressing this problem, and their preliminary results are encouraging. This area of research will continue to be a major focus in organic chemistry, and future strategies for trifluoromethylation will certainly be influenced by the current reports[3,4].

In the shorter term, these breakthroughs open the door to the preparation of a diverse range of molecules that would previously have been difficult to access. The ability of these streamlined approaches to perform late-stage trifluoromethylations on structurally complex molecules will undoubtedly have an impact on the work of research chemists in the pharmaceutical, agrochemical and other industries. ■

Andrew T. Parsons *and* **Stephen L. Buchwald**
are in the Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
e-mails: sbuchwal@mit.edu; parsonsa@mit.edu

1. Muller, K., Faeh, C. & Diederich, F. *Science* **317,** 1881–1886 (2007).
2. Hiyama, T. *Organofluorine Compounds: Chemistry and Applications* (Springer, 2000).
3. Ji, Y. *et al. Proc. Natl Acad. Sci. USA* **108,** 14411–14415 (2011).
4. Nagib, D. A. & MacMillan, D. W. C. *Nature* **480,** 224–228 (2011).
5. Tomashenko, O. A. & Grushin, V. V. *Chem. Rev.* **111,** 4475–4521 (2011).
6. Langlois, B. R., Laurent, E. & Roidot, N. *Tetrahedron Lett.* **32,** 7525–7528 (1991).
7. Kamigata, N., Fukushima, T. & Yoshida, M. *Chem. Lett.* **19,** 649–650 (1990).
8. Kamigata, N., Ohtsuka, T., Fukushima, T., Yoshida, M. & Shimizu, T. *J. Chem. Soc. Perkin Trans. 1* 1339–1346 (1994).

S.L.B. declares competing financial interests.
See go.nature.com/rdc42q for details.

STEM CELLS

The clock within

The molecular clock machinery regulates organisms' responses to daily variations in the environment. One unexpected response seems to be temporal fine-tuning of stem-cell behaviour in the skin. SEE ARTICLE P.209

LORENA AGUILAR-ARNAL & PAOLO SASSONE-CORSI

In general, the physiology and behaviour of most organisms follow a daily, or circadian, rhythm. This contributes to adaptations to environmental variables such as the light–dark cycle. Although in mammals the coherent ticking of circadian oscillations is ultimately orchestrated by a tiny region of the brain — the suprachiasmatic nucleus[1] — the oscillations are intrinsic to almost all tissues and cells, including the epidermal layers of the skin. Circadian oscillations in gene expression are controlled by a number of factors that constitute the molecular clock. Strikingly, in the epidermis this molecular clock is not solely dedicated to the regulation of the tissue's circadian rhythms. On page 209 of this issue, Janich *et al.*[2] report that the clock also controls the alternate cycles of dormancy and proliferation in mouse epidermal stem-cell populations. This finding places the molecular clock in a strategic position for the fine-tuning of epidermal homeostasis.

Adult stem cells are present in various organs and tissues, where they reside in distinct areas within specialized microenvironments, or 'niches'. They remain in a slow-cycling state until needed for the maintenance of tissue homeostasis. However, they can also become active in response to malignant signals, leading to inflammation and cancer. The transition of stem cells from dormancy to proliferation remains poorly understood. One question is why only a few stem cells of a population respond to proliferation and differentiation signals. Janich and colleagues provide unanticipated insight into this aspect of stem-cell biology.

The molecular machinery controlling circadian functions involves a complex network

© 2011 Macmillan Publishers Limited. All rights reserved

synthetic routes, without first preparing pre-functionalized or trifluoromethyl-containing substrates. It would therefore decrease the number of synthetic operations necessary to access a variety of trifluoromethylated molecules (Fig. 1c).

The authors[3,4] recognized that the development of a successful method hinged on identifying a practical trifluoromethyl source. Both groups independently surmised that the reliable generation of highly reactive trifluoromethyl radicals ($^{\bullet}CF_3$) as intermediates might allow the direct trifluoromethylation of non-functionalized substrates, in what Ji *et al.*[3] call an "innate trifluoromethylation". In other words, the inherent reactivity of the substrates would dictate the site (or sites) at which the trifluoromethylation reaction occurs.

The two groups pursued different approaches to develop their own mild, efficient methods to access reactive trifluoromethyl-radical intermediates. Ji *et al.* developed a transition-metal-free approach that relies solely on air- and moisture-stable reagents (which are easier to handle than commonly used organometallic reagents and transition-metal catalysts that react with air and moisture). They determined that the Langlois reagent[6] ($NaSO_2CF_3$) serves as a reliable source of $^{\bullet}CF_3$ when treated with *t*-butylhydroperoxide (a commonly used reagent in radical chemistry). Nagib and MacMillan[4], however, chose triflyl chloride ($ClSO_2CF_3$) as their source of $^{\bullet}CF_3$ (refs 7, 8). They generated the radicals simply by shining light from a household fluorescent bulb onto a solution of triflyl chloride in the presence of a light-activated ruthenium catalyst. Both methods[3,4] generated $^{\bullet}CF_3$ in the presence of a diverse range of substrates, and provided access to an array of structurally complex trifluoromethylated molecules. Furthermore, both methods worked reliably even for substrates bearing potentially sensitive functional groups (those whose presence in a molecule often adversely affects the course of a range of chemical reactions, including traditional trifluoromethylation reactions).

To demonstrate the versatility of their methods, both groups conducted trifluoromethylations of common pharmaceuticals and natural products. Ji *et al.*[3] successfully achieved C–H trifluoromethylations of the natural-product derivative dihydroquinine and the billion-dollar drug Chantix (varenicline, used to combat nicotine addiction), providing products bearing a single trifluoromethyl group. These examples are significant because accessing CF_3–dihydroquinine and CF_3–Chantix using previously available trifluoromethylation methods would require laborious, multi-step total syntheses. Notably, the reactions were chemoselective, even though the starting materials contained multiple reactive sites.

Similarly, Nagib and MacMillan[4] subjected the cholesterol-lowering drug Lipitor (atorvastatin) to their reaction conditions. This experiment resulted in the non-selective formation of three chromatographically separable isomers of CF_3–Lipitor, in which the trifluoromethyl group was appended to different carbons in the molecule. The improved efficiency of this approach for accessing new chemical entities cannot be overstated, given that preparation of these isomers using traditional methods would require three parallel total syntheses, which would be extremely time- and labour-intensive.

Although the new methods[3,4] constitute a major contribution to the field of trifluoromethylation chemistry, many challenges remain. For example, it would be desirable to have precise control over the site at which C–H trifluoromethylation occurs. This might be accomplished through the development of reaction conditions that combine the best attributes of the currently available methods. Complementary protocols that selectively produce different trifluoromethyl-containing molecules from a single substrate would also be advantageous, because it would avoid the often difficult task of separating mixtures of isomers generated as products. Ji *et al.*[3] report their initial efforts towards addressing this problem, and their preliminary results are encouraging. This area of research will continue to be a major focus in organic chemistry, and future strategies for trifluoromethylation will certainly be influenced by the current reports[3,4].

In the shorter term, these breakthroughs open the door to the preparation of a diverse range of molecules that would previously have been difficult to access. The ability of these streamlined approaches to perform late-stage trifluoromethylations on structurally complex molecules will undoubtedly have an impact on the work of research chemists in the pharmaceutical, agrochemical and other industries. ■

Andrew T. Parsons *and* **Stephen L. Buchwald**
are in the Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
e-mails: sbuchwal@mit.edu; parsonsa@mit.edu

1. Muller, K., Faeh, C. & Diederich, F. *Science* **317,** 1881–1886 (2007).
2. Hiyama, T. *Organofluorine Compounds: Chemistry and Applications* (Springer, 2000).
3. Ji, Y. *et al. Proc. Natl Acad. Sci. USA* **108,** 14411–14415 (2011).
4. Nagib, D. A. & MacMillan, D. W. C. *Nature* **480,** 224–228 (2011).
5. Tomashenko, O. A. & Grushin, V. V. *Chem. Rev.* **111,** 4475–4521 (2011).
6. Langlois, B. R., Laurent, E. & Roidot, N. *Tetrahedron Lett.* **32,** 7525–7528 (1991).
7. Kamigata, N., Fukushima, T. & Yoshida, M. *Chem. Lett.* **19,** 649–650 (1990).
8. Kamigata, N., Ohtsuka, T., Fukushima, T., Yoshida, M. & Shimizu, T. *J. Chem. Soc. Perkin Trans. 1* 1339–1346 (1994).

S.L.B. declares competing financial interests.
See go.nature.com/rdc42q for details.

STEM CELLS

The clock within

The molecular clock machinery regulates organisms' responses to daily variations in the environment. One unexpected response seems to be temporal fine-tuning of stem-cell behaviour in the skin. SEE ARTICLE P.209

LORENA AGUILAR-ARNAL & PAOLO SASSONE-CORSI

In general, the physiology and behaviour of most organisms follow a daily, or circadian, rhythm. This contributes to adaptations to environmental variables such as the light–dark cycle. Although in mammals the coherent ticking of circadian oscillations is ultimately orchestrated by a tiny region of the brain — the suprachiasmatic nucleus[1] — the oscillations are intrinsic to almost all tissues and cells, including the epidermal layers of the skin. Circadian oscillations in gene expression are controlled by a number of factors that constitute the molecular clock. Strikingly, in the epidermis this molecular clock is not solely dedicated to the regulation of the tissue's circadian rhythms. On page 209 of this issue, Janich *et al.*[2] report that the clock also controls the alternate cycles of dormancy and proliferation in mouse epidermal stem-cell populations. This finding places the molecular clock in a strategic position for the fine-tuning of epidermal homeostasis.

Adult stem cells are present in various organs and tissues, where they reside in distinct areas within specialized microenvironments, or 'niches'. They remain in a slow-cycling state until needed for the maintenance of tissue homeostasis. However, they can also become active in response to malignant signals, leading to inflammation and cancer. The transition of stem cells from dormancy to proliferation remains poorly understood. One question is why only a few stem cells of a population respond to proliferation and differentiation signals. Janich and colleagues provide unanticipated insight into this aspect of stem-cell biology.

The molecular machinery controlling circadian functions involves a complex network

© 2011 Macmillan Publishers Limited. All rights reserved

of auto-regulatory feedback loops that affect gene expression. The positive components of this machinery — the transcriptional activators CLOCK and BMAL1 — drive rhythmic expression of the clock-controlled genes. Among these are the genes that encode the negative components of the molecular clock, the repressor proteins cryptochrome (CRY1 and 2) and period (PER1–3)[3]. Additional factors, such as other clock-related proteins, post-translational modifications and small molecules that can feed back into the clock, add further complexity to the system, thereby ensuring regulated expression of at least 10% of the cellular genes in a circadian manner[3].

Janich *et al.*[2] studied a mouse model engineered to express a fluorescent protein in a way that reflects circadian clock oscillations. The authors find that, in the hair-follicle bulge (one of the two locations where epidermal stem cells reside), two subpopulations of stem cells coexist. These differ in the way that they express the core components of the clock machinery (Fig. 1). This, in turn, results in differential expression of several genes known to control stem-cell dormancy, proliferation and differentiation. Among these are genes that encode TGF-β factors, which inhibit proliferation and differentiation, and Wnt-signalling factors, which promote growth. Consequently, at any one time, one subpopulation is more prone to proliferation and differentiation.

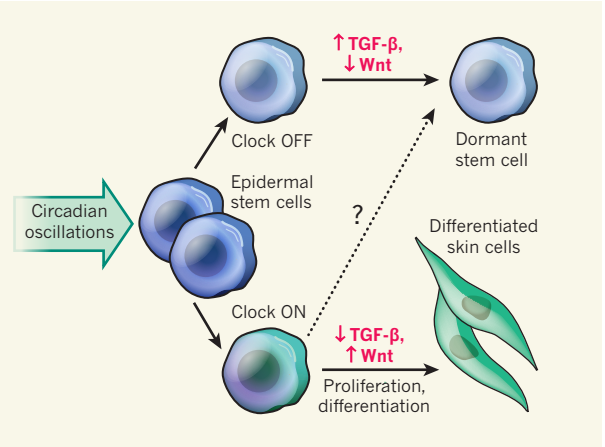


Figure 1 | Stem cells and circadian oscillations. Epidermal stem cells coexist as two subpopulations: one that remains dormant and the other that becomes active to maintain epidermal homeostasis. Janich *et al.*[2] report that, in response to circadian oscillations, the two subpopulations differentially express components of the clock machinery. This, in turn, affects expression of proteins such as Wnt signalling factors, which promote proliferation, and TGF-β factors, which inhibit growth. Other factors such as metabolites and hormones could also be involved in regulating epidermal stem-cell heterogeneity. The 'decision-making' processes probably involve metabolic and epigenetic control. Whether activated stem cells can become dormant again remains unknown.

What could be the advantage of two stem-cell subpopulations coexisting in the same niche? Stem-cell heterogeneity may have evolved to allow the cells both to self renew, thus replenishing their reserve in the niche, and to keep a ready-to-go population that can respond to the signals that trigger differentiation. Whether the two epidermal stem-cell subpopulations represent two lineages with distinct developmental origins, and whether they can interconvert, should be determined (Fig. 1).

In subsequent experiments, Janich and co-workers disrupted circadian rhythm in mice by selectively deleting the *Bmal1* gene in the basal keratinocyte cells of the animals' skin. The hair-follicle bulges of these mice contained fewer proliferative cells and a higher number of dormant stem cells than those of normal mice. Moreover, *Bmal1*-deficient bulge stem cells were more responsive to TGF-β signalling and exhibited decreased expression of Wnt-related genes while maintaining higher expression levels of TGF-β-related genes. When the authors deleted the clock inhibitors *Per1* and *Per2*, they observed the opposite effect — increased proliferation of bulge stem cells. Thus, the positive and negative components of the clock machinery seem to function as alternative rheostats of epidermal stem-cell proliferation.

Previous studies have also suggested a role for the circadian clock in stem-cell biology. For example, the release of haematopoietic stem cells — which develop into blood cells — from the bone marrow into the bloodstream follows a cyclical pattern that seems to be regulated by the circadian clock through the sympathetic nervous system[4].

Adult stem-cell differentiation, which is essential for tissue homeostasis, has also been linked to circadian oscillations. Clock genes, for instance, mediate inhibition of bone formation from osteoblast cells, and several mouse models deficient in clock genes exhibit increased bone mass relative to controls[5]. Moreover, embryonic fibroblasts derived from mice lacking either *Bmal1* or *Per2* show altered ability to differentiate into adipocytes[6,7], indicating that the clock machinery affects formation of the fat tissue. Janich *et al.* also find that disruption of *Bmal1* in the skin of mice induces epidermal ageing and reduces tumorigenesis. These pieces of evidence clearly point to a role for the clock in maintaining skin-tissue homeostasis.

Janich and co-authors' findings[2] consolidate the existing idea of crosstalk between circadian rhythms and differentiation. They also raise several questions. The undifferentiated state of the cell has been correlated with a lack of circadian physiology[8]. It is therefore important to determine which stimuli initiate the clock's ticking in the niche. It could be that, whereas differentiating stimuli can mediate this process in embryonic stem cells[8], signals originating from the central nervous system activate adult stem cells. Other unknowns include how and when communication between the central pacemaker in the suprachiasmatic nucleus and the peripheral clocks it synchronizes is established, and how relevant it is to the differentiation process.

An emerging theme is that the metabolic state of cells is crucial for clock regulation and that the clock machinery contributes to the maintenance of the metabolic state. Indeed, the finding that the metabolic mediator NAD^+ and its biosynthetic pathway are components of the circadian oscillatory network[9,10] is evidence for modulation of the clock machinery through metabolic pathways. Because stem-cell metabolism is reprogrammed in response to differentiation signals[11], the possibility that cyclical metabolism may contribute to epidermal stem-cell physiology is intriguing. Deciphering the relationship between circadian oscillations and stem-cell metabolism could lead to new targets to manipulate the differentiation and function of adult stem cells for regenerative medicine.

The information conveyed by differentiation signals is integrated into stem cells epigenetically — through chemical modifications of the genome that do not affect the DNA sequence but are inherited after cell division — to establish the cells' fate. Does the clock machinery contribute to this process? Accumulating evidence indicates that intricate chromatin (DNA–protein complexes) remodelling events that are involved in the establishment of several epigenetic marks in the genome follow a circadian pattern, and ultimately control circadian gene expression[12]. Establishment of such a clock-regulated epigenome during differentiation raises the attractive possibility that differentiating cells use the clock machinery to direct the epigenetic marks that define certain gene-transcription patterns in the new cell type. The existence of these fascinating links should become clear, as epigenetic features of epidermal stem cells are already being explored[13,14].

This paper[2] raises yet another intriguing question: is this mechanism of clock-mediated epidermal stem-cell control conserved in other adult stem cells? Expect exciting insight into stem-cell biology in the light of Janich and co-workers' data. ■

Lorena Aguilar-Arnal *and* **Paolo Sassone-Corsi** *are at the Center for Epigenetics and Metabolism, School of Medicine, University of California, Irvine, Irvine, California 92697-4625, USA.*
e-mail: psc@uci.edu

1. Schibler, U. & Sassone-Corsi, P. *Cell* **111,** 919–922 (2002).
2. Janich, P. *et al. Nature* **480,** 209–214 (2011).

© 2011 Macmillan Publishers Limited. All rights reserved

3. Zhang, E. E. & Kay, S. A. *Nature Rev. Mol. Cell Biol.* **11,** 764–776 (2010).
4. Méndez-Ferrer, S., Lucas, D., Battista, M. & Frenette, P. S. *Nature* **452,** 442–447 (2008).
5. Fu, L., Patel, M. S., Bradley, A., Wagner, E. F. & Karsenty, G. *Cell* **122,** 803–815 (2005).
6. Shimba, S. *et al. Proc. Natl Acad. Sci. USA* **102,** 12071–12076 (2005).
7. Grimaldi, B. *et al. Cell Metab.* **12,** 509–520 (2010).
8. Yagita, K. *et al. Proc. Natl Acad. Sci. USA* **107,** 3846–3851 (2010).
9. Nakahata, Y., Sahar, S., Astarita, G., Kaluzova, M. & Sassone-Corsi, P. *Science* **324,** 654–657 (2009).
10. Ramsey, K. M. *et al. Science* **324,** 651–654 (2009).
11. Nakada, D., Levi, B. P. & Morrison, S. J. *Neuron* **70,** 703–718 (2011).
12. Masri, S. & Sassone-Corsi, P. *Nature Neurosci.* **13,** 1324–1329 (2010).
13. Luis, N. M. *et al. Cell Stem Cell* **9,** 233–246 (2011).
14. Mejetta, S. *et al. EMBO J.* **30,** 3635–3646 (2011).

ASTROPHYSICS

Monster black holes

A combination of ground-based and spacecraft observations has uncovered two black holes of 10 billion solar masses in the nearby Universe. The finding sheds light on how these cosmic monsters co-evolve with galaxies. SEE LETTER P.215

MICHELE CAPPELLARI

Giant black holes, with masses of a few billion times that of the Sun, have fascinated scientists, science-fiction writers and the general public alike since they were first proposed[1] four decades ago. These supermassive black holes, which lurk in the centre of galaxies, are not just theoretical curiosities. Their existence was convincingly demonstrated 15 years ago, and soon astrophysicists realized that such black holes can have profound effects on how galaxies form[2]. This realization spurred a flurry of studies aiming to understand the joint evolution of galaxies and black holes. On page 215 of this issue, McConnell *et al.*[3] present the detection of the two most massive black holes ever found. They argue that their finding provides a key missing piece of evidence to our understanding of how galaxies and black holes form.

The existence of monster black holes was originally invoked to explain the intense energy released by active galaxies known as quasars. These galaxies are especially numerous at very large distances from Earth, corresponding to a time when the Universe was less than half its current age. Their tremendous brightness is thought to be the last flash from gas accelerated to extreme speed before being accreted by a giant black hole in the galaxy's nucleus. There is now less gas in the nearby Universe than there was at early epochs, because most of it was used to make stars. For this reason, no quasar exists in our cosmic backyard. However, if supermassive black holes are responsible for quasars' power production, they should still lie dormant in the centre of the most massive nearby galaxies.

Indeed, systematic searches for supermassive black holes have found them at the centre of all massive galaxies for which reliable determinations of the black-hole mass exist (see McConnell *et al.*[3] for an up-to-date list). But such black holes are not massive enough to power the brightest quasars. The ideal candidate galaxies for hosting these largest black holes are the massive elliptical galaxies found at the centre of galaxy clusters. Like giant spiders at the centre of a web, these galaxies lie at the bottom of a cluster's gravitational potential well, feeding their black hole by accreting gas and stars from neighbouring galaxies. However, the nearest big galaxy cluster to Earth is about 100 megaparsecs away, and this large distance makes determining the mass of their putative black holes a challenge.

McConnell *et al.*[3] targeted the central elliptical galaxy of two massive clusters, using two of the world's largest telescopes and observations from the Hubble Space Telescope. They found two black holes, each with a mass of more than 10 billion solar masses, in the nuclei of the two galaxies. These objects probably represent the missing dormant relics of the giant black holes that powered the brightest quasars in the early Universe.

The mass of the supermassive black holes studied so far is closely related to the amount of random motion — the velocity dispersion — of the stars in the central parts of the host galaxies[4,5]. The velocity dispersion is related to the mass of the spheroidal component, the stellar bulge, of the galaxies. The existence of this empirical relationship is attributed to the fact that the black hole grows by feeding from the gas that is accreted by its host galaxy. This gas accretion leads to the simultaneous growth of both the black hole and the galaxy stellar bulge, until the intense energy produced by the black hole heats up or blows away the gas, halting both star formation and black-hole accretion[6] (Fig. 1a).

Interestingly, the two newly measured supermassive black holes are more massive than would be predicted from their velocity dispersion. This suggests that, unlike their smaller counterparts, these black holes did not grow most of their mass by gas accretion but instead grew by the 'dry' merging of gas-poor galaxies. In a dry merger of two similar galaxies, the resulting galaxy's stellar and black-hole masses are the sums of the corresponding masses of the two progenitor galaxies (Fig. 1b). However,

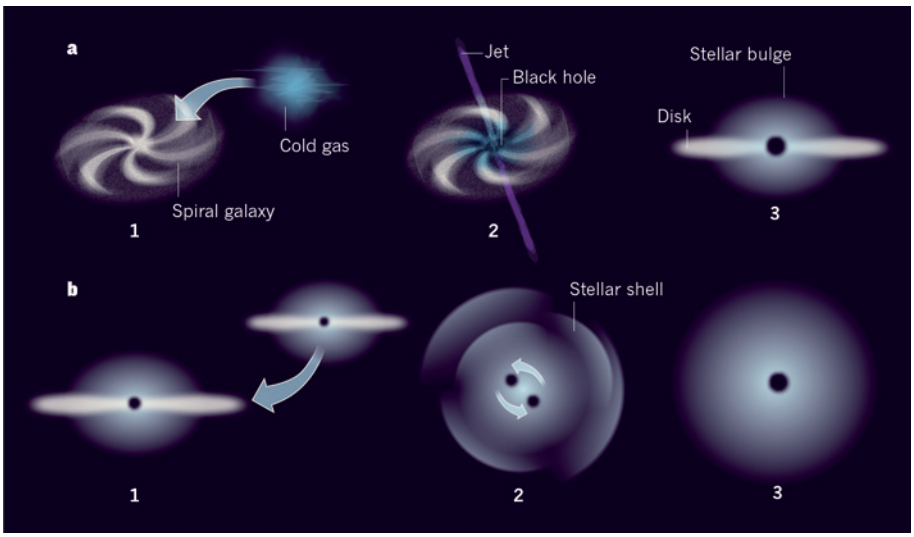


Figure 1 | Two ways to grow galaxies and their central black holes. **a**, In the gas-accretion hypothesis, a pure-disk, gas-rich spiral galaxy, possibly with a small black hole, accretes cold gas either from the environment (1) or by merging with another spiral galaxy (not shown). The gas sinks to the galaxy's centre, forming a black hole and a stellar bulge. This process terminates when the galaxy becomes 'active' and produces two opposing high-speed jets of particles (2). The end product is a lenticular galaxy, which has a supermassive black hole, disk and stellar bulge, and larger stellar velocity dispersion than the initial galaxy (3). **b**, Alternatively, the galaxy and black hole can grow by 'dry' merging: a gas-poor lenticular galaxy accretes a similar galaxy (1). The two galaxies merge and form stellar shells, while their black holes spiral towards the centre until they coalesce (2). The result is a spheroidal galaxy that has a larger size and bigger black hole than the progenitors, but nearly unchanged velocity dispersion (3). McConnell and colleagues' analysis[3] indicates that the second route may dominate the growth of the most massive black holes.

© 2011 Macmillan Publishers Limited. All rights reserved

contrary to the gas-accretion case, the velocity dispersion of the final galaxy's stars remains unchanged, or changes little if the mass doubling is achieved through a sequence of several smaller mergers of gas-poor galaxies[7]. Either way, the merging process leaves behind a black hole that is more massive than would be predicted from the mass–velocity dispersion relationship — just like the two black holes observed by McConnell and colleagues[3].

Not everybody agrees that the observed mass–velocity dispersion relationship indicates that black holes regulate galaxy growth by stopping gas accretion with their energy production (Fig. 1a). Studies[8,9] have shown that starting from a set of galaxies containing black holes of random masses, a sequence of mergers can lead to a relationship close to the observed one without the need for energy input from the black hole. However, this simulated relationship deviates systematically from the observed one. These claims[8,9] can be tested by accurately measuring black-hole masses in galaxies with different properties, as McConnell *et al.* did. Progress requires the largest telescopes and the use of state-of-the-art technologies such as integral-field spectroscopy, to map the motion of the stars in two dimensions, and adaptive optics, to correct the blurring effect of Earth's atmosphere and achieve sharp images[10]. The future looks bright for black-hole studies using the next generation of 40-metre telescopes, such as the European Extremely Large Telescope, which will significantly increase the number of galaxies that can be reliably investigated. ■

Michele Cappellari *is in the Sub-department of Astrophysics, Department of Physics, University of Oxford, Oxford OX1 3RH, UK.*
e-mail: cappellari@astro.ox.ac.uk

1. Lynden-Bell, D. *Nature* **223,** 690–694 (1969).
2. Silk, J. & Rees, M. J. *Astron. Astrophys.* **331,** L1–L4 (1998).
3. McConnell, N. J. *et al. Nature* **480,** 215–218 (2011).
4. Ferrarese, L. & Merritt, D. *Astrophys. J.* **539,** L9–L12 (2000).
5. Gebhardt, K. *et al. Astrophys. J.* **539,** L13–L16 (2000).
6. Di Matteo, T., Springel, V. & Hernquist, L. *Nature* **433,** 604–607 (2005).
7. Naab, T., Johansson, P. H. & Ostriker, J. P. *Astrophys. J.* **699,** L178–L182 (2009).
8. Peng, C. Y. *Astrophys. J.* **671,** 1098–1107 (2007).
9. Jahnke, K. & Macció, A. V. *Astrophys. J.* **734,** 92 (2011).
10. Cappellari, M. *et al. Mon. Not. R. Astron. Soc.* **394,** 660–674 (2009).

GENETICS

Noise rules

The idea that gene variants alone control an organism's traits is overly simple. A study of the effects of gene interactions on the outcomes of random variation in gene expression reveals the complex reality. SEE LETTER P.250

HANA EL-SAMAD & JONATHAN S. WEISSMAN

Ever since Gregor Mendel and Thomas Morgan crystallized the idea of the gene as the heritable unit, biologists have been trying to map the relationship between genotype (the complement of gene variants present in an individual) and phenotype (an individual's physical characteristics). On page 250 of this issue, Lehner and colleagues[1] provide insight into the complexity of this relationship by showing how random variation in the expression of members of pairs of interacting genes can act synergistically to affect the development of a complex organism — the roundworm *Caenorhabditis elegans*.

Notable advances in DNA sequencing during the past few years have made it possible to define the genotype of any individual rapidly and cheaply. As a result, scientists glibly talk about the possibility of pinpointing individual genes that influence every aspect of our being, from our propensity to get cancer to our chances of living to be 100 years old. However, this appealing vision oversimplifies a much more complex reality. There is no one-to-one relationship between genotype and phenotype — even identical twins can have radically different personalities, disease susceptibilities and life trajectories.

Although many factors contribute to the mapping of genotype to phenotype, the influence of two of these is especially challenging to account for. The first is 'noise' — the molecular fluctuations inherent in biological systems that cause random switching of genes on and off. The second is that genes do not act alone, but work together to form functional cellular networks. That results in complex genetic interactions in which the phenotypic impact of a particular variant (or allele) of a gene is dependent on an individual's full genetic make-up. For example, it is not uncommon for individual alleles to be beneficial on their own, but lethal when two copies are present.

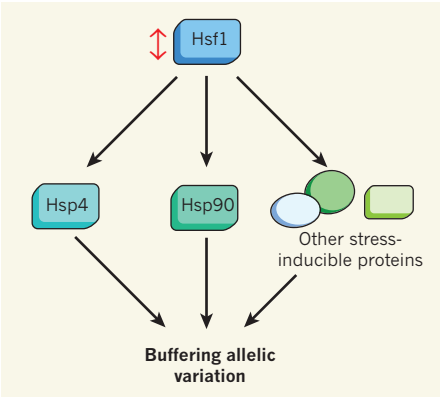


Figure 1 | Noise regulons and gene expression. Noise regulons are synchronized stochastic changes in functionally related genes, and are coordinated by an upstream signal. In this example, fluctuations (red arrow) in the expression of the transcription factor Hsf1 cause the coordinated expression of stress-response proteins, including the molecular chaperones Hsp4 and Hsp90. These proteins then cooperate to buffer the effects of a mutant allele. Such a mechanism helps explain why loss of *tbx-9* in *C. elegans* causes abnormal development in only half of the worms affected by this mutation, a phenomenon studied by Lehner and colleagues[1].

Alone, either of these factors could generate an elaborate relationship between genotype and phenotype, but there is inevitably a complex interplay between the two. Lehner and colleagues' study[1] examines this interplay. Previous work[2] has investigated how noise in gene expression influences the effects of a defective allele of the *C. elegans skn-1* gene. Notably, the *skn-1* mutant allele leads to defects in intestinal development in some, but not all, individuals; such incomplete development of a phenotypic trait is known as partial penetrance.

Using a clever approach (known as fluorescence *in situ* hybridization) that enables the exact number of RNA molecules in a cell to be counted, this study[2] established that random fluctuations in certain factors lead to a variable 'all-or-none' expression of the transcription factor gene *elt-2*, which is dependent on *skn-1*. In wild-type worms, the system is robustly buffered against such fluctuations. But in the *skn-1* mutants, the variability is unleashed, so that *elt-2* is not activated in some cells, hence preventing normal development of the gut. These results illustrate how quantitative fluctuations in upstream components of a signal-transduction system can lead to qualitative differences in a downstream developmental event.

Lehner and colleagues[1] further explore the impact of stochastic fluctuations on the partial penetrance of a mutation, while also examining their interplay with the gene network in which the mutant allele is involved. In particular, the authors address the question of how fluctuations in factors expressed early in development affect subsequent developmental switches. This necessitated the development of a new experimental strategy, because in the previous work[2] the worms were killed to measure messenger RNA levels, thereby precluding the observation of later developmental events.

© 2011 Macmillan Publishers Limited. All rights reserved

The authors[1] used transcriptional reporters — in which a second copy of the gene of interest drives the expression of a fluorescent protein — to follow variations in gene expression in live worms. A key requirement of this approach (verified by the authors) is that fluctuations in expression of the reporter accurately reflect fluctuations in expression of the endogenous genes. This means that the reporter is tracking expression noise resulting from differences in cellular environment — extrinsic noise[3] — rather than the random turning on and off of individual promoters.

Lehner and co-workers first investigated why mutations in the T-box transcription factor gene *tbx-9* that completely abolish the gene's function cause an incompletely penetrant defect in *C. elegans* larval development: roughly half of the animals lacking *tbx-9* develop normally; the other half have muscle and epidermal defects. The authors observed that overexpression of *tbx-8* — a gene closely related to *tbx-9* — eliminates the defects caused by loss of *tbx-9*. What's more, loss of *tbx-9* in their mutants caused upregulation of *tbx-8*.

More strikingly, the authors observed that differences in *tbx-8* expression were a strong predictor of the phenotypic effects of *tbx-9* loss — *tbx-9*-mutant worms that expressed *tbx-8* at a low level were considerably more likely to develop abnormally than were genotypically identical individuals that expressed *tbx-8* more strongly. Thus, stochastic differences in the feedback loop that compensates for *tbx-9* loss by upregulating *tbx-8* contribute to the variable penetrance of the *tbx-9* mutation. Lehner and colleagues observed a similar set of effects with the transcription factors *flh-1* and *flh-2*, which, like *tbx-8* and *tbx-9*, are a pair of related genes that resulted from ancient gene duplication. This type of noisy feedback between related genes may therefore be quite common.

Although these findings represent a step forwards, the variability of *tbx-8* expression could not fully account for the variable penetrance of the *tbx-9* mutation. What, then, is the source of the remaining variability? Molecular chaperones — proteins that assist other proteins in folding — can buffer a wide range of cellular defects, especially the Hsp90 chaperone[4] (also known as DAF-21 in *C. elegans*). Lehner and co-workers therefore investigated random fluctuations in chaperone levels in individual worms as a possible explanation of the missing variability. They found that fluctuations in expression of *daf-21* were a strong predictor of the effects of *tbx-9* loss. Importantly, the effects of the differences in *tbx-8* and *daf-21* expression were independent of each other but synergistic: more than 90% of worms with high levels of *daf-21* and *tbx-8* expression developed normally, whereas roughly two-thirds of the animals in which expression of both genes was low did not.

Lehner and colleagues found that the observed variability in penetrance of *tbx-9* loss in their study can, to a remarkable degree, be accounted for by variations in expression of *tbx-8* and *daf-21*. This, in turn, raises the question of what underlying mechanisms in the cell cause the observed variability and how many other genes are affected by these fluctuations. In the case of *daf-21*, the authors show that the fluctuations are probably part of coordinated changes in the expression of a broad range of chaperone genes, including one known as *hsp-4*. Such 'noise regulons' — synchronized stochastic changes in functionally related genes — are particularly suitable for causing coordinated effects on cell physiology, because the related genes act together to carry out common functions[5] (Fig. 1). However, the authors did not find evidence that fluctuations in *tbx-8* expression correlate with variations in the expression of chaperones, and it may be that there are no other genes whose noise co-varies with that of *tbx-8*. A more tantalizing possibility is that variations in the expression of *tbx-8* also reflect a coherent cellular state in which other members of the *tbx-8*/*tbx-9* pathway fluctuate in unison. This raises the further question of how many such noise regulons exist, and what controls their activity.

More broadly, it is clear from recent studies[1,2] that stochastic fluctuations can have a big impact on the penetrance of gene alleles. A better understanding of the structure of noise — which genes tend to fluctuate together, and how these fluctuations are controlled — should provide crucial insight into the nature of the genotype–phenotype relationship. ■

Hana El-Samad *is in the Department of Biochemistry and Biophysics, and* **Jonathan S. Weissman** *is in the Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, California 94158, USA. J.S.W. is also at the Howard Hughes Medical Institute.*
e-mails: hana.el-samad@ucsf.edu; weissman@cmp.ucsf.edu

1. Burga, A., Casanueva, M. O. & Lehner, B. *Nature* **480,** 250–253 (2011).
2. Raj, A., Rifkin, S. A., Andersen, E. & van Oudenaarden, A. *Nature* **463,** 913–918 (2010).
3. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. *Science* **297,** 1183–1186 (2002).
4. Rutherford, S. L. & Lindquist, S. *Nature* **396,** 336–342 (1998).
5. Stewart-Ornstein, J., Weissman, J. S. & El-Samad, H. *Mol. Cell* (in the press).

MATERIALS SCIENCE

A sense for touch

Will a sense of touch similar to that of humans ever be developed in robots? Results on the physics of friction for fingerprint-like ridges sliding across textured surfaces may lead the way to tactile robotic sensors.

C. MATHEW MATE & ROBERT W. CARPICK

When we rub a finger across a surface, our sense of touch is remarkably adept at distinguishing between different textures[1]. For example, astute clothes shoppers can easily feel the difference in texture between cotton and lower quality polyester fabrics, just as experienced cashiers can spot counterfeit banknotes from the feel of the paper. Writing in *Physical Review Letters*, Wandersman *et al.*[2] provide a major advance towards understanding the physics behind these tactile sensations, showing how a modulated frictional signal is generated by fingerprint-like ridges rubbing against the roughness of an opposing surface. This improved understanding of the relationship between friction and surface textures should have an impact in many areas, but particularly in the field of tactile sensors for robotic design. Incorporating sensors capable of mimicking the human sense of touch has long been recognized[3] as important for improving the ability of robots to grasp objects firmly without damaging them.

Wandersman *et al.*[2] present measurements and analysis of the tangential or friction force generated as a rubbery elastomer block with periodic surface ridges slides over a rough glass surface. The ridges on the elastomer surface, which have periods ranging from 125 to 760 micrometres, serve as stand-ins for the epidermal ridges on human fingers (Fig. 1a). They show that, when pressed against the glass surface to form contact regions several millimetres in diameter, the elastomer deforms around the rough surface texture in the same way as the epidermal ridges would if pressed against such a surface.

With numerous ridges in contact, one might expect that any periodic variation in the friction force due to these ridges would average out; indeed, this can be shown analytically if the friction force generated in each microscopically small area in contact (much smaller than the width of a ridge) is strictly proportional to the local normal or loading force pressing the two surfaces into contact over this area. This linear relationship between the friction and load forces is usually referred to as Amontons' law of friction[4], which states that the friction force, F, is proportional to

© 2011 Macmillan Publishers Limited. All rights reserved

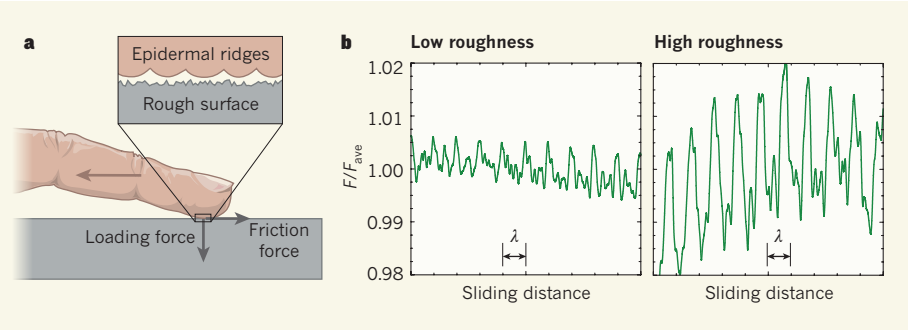


Figure 1 | The friction force. **a**, When a fingertip is rubbed against a rough surface, the net friction force acting on the epidermal ridges increases with the loading force that the finger exerts on the ridges to press them into contact with the surface. **b**, Wandersman *et al.*[2] measure the net friction force that acts on an elastomer block as it slides against glass surfaces of differing roughness. The elastomer block has ridges similar in structure and elasticity to the epidermal ridges on a finger. Shown here is the instantaneous friction force F, normalized to the average friction force F_{ave}, as a function of sliding distance. λ is the spacing of the ridges on the elastomer block and is 218 micrometres. The force has a slight oscillating component that has the same period as the separation between the ridges and increases with the degree of roughness on the glass. (Part **b** modified from ref. 2.)

the loading force, L, or $F = \mu L$, where μ is the coefficient of friction. However, the authors find that, rather than this friction modulation averaging to zero, the net friction acting on the elastomer block goes up and down slightly (by up to a few per cent) during the sliding experiments (Fig. 1b), with the same period as the ridges. And, surprisingly, the amplitude of this modulation actually increases when the net loading force increases.

One of the fundamental advances provided by Wandersman *et al.*[2] is their analysis showing that, if the local friction coefficient depends even slightly on pressure (which is equivalent to friction being slightly non-linear with load), the modulation in friction can increase with loading force. They find good agreement between the experiments and a model in which friction varies non-linearly with load: $F = AL^{\gamma}$, where $\gamma = 0.87 \pm 0.04$ and A is a constant. In this model, the roughness of the surface against which the fingerprint-like ridges are being slid has the important role of providing a heterogeneous distribution of contact pressure locally along the ridges. As the roughness increases, a wider distribution of loading pressure occurs, leading to a larger modulation in friction as a result of the non-linear nature of friction with load. The spatial period of the ridges serves to concentrate the minute variation of friction caused by these texture-induced pressure modulations all at one frequency, making it much easier to discern this variation from the average net friction.

The sliding of fingerprint-like ridges over surfaces is not the only area in which Wandersman and colleagues' analysis should apply. Because friction forces are rarely strictly linear with loading forces (as postulated in Amontons' law)[5], we believe that this analysis could provide a valuable way to use friction fluctuations to characterize surface roughness on many types of material pairs sliding against each other. The amplitude and the load-dependence of the fluctuations reveal information on the surface's topographic characteristics at length scales much smaller than that of the patterned ridges. As a result, we think that one exciting area to which the method developed by Wandersman *et al.*[2] could be extended is the characterization of surface roughness down to the nanometre scale or even smaller atomic length scales.

For example, for many years, atomic-scale modulations of friction have been observed when the sharp tip of an atomic force microscope (AFM) slides across the periodic arrangement of atoms on a crystalline surface[6]. However, these AFM experiments typically require very small loading forces (nanonewtons) to maintain a contact area of only a few nanometres in diameter in order to see the atomic-scale modulation of friction. But, perhaps, with suitably designed patterned ridges and friction sensors, this ability to sense the atomic-level contribution to the friction modulation could be extended from the current nanometre-sized contact zones of AFMs to millimetre-sized contact zones, allowing future robotic fingers to feel the atomic-level contribution to surface texture. ■

C. Mathew Mate *is at the Hitachi San Jose Research Center, San Jose, California 95135, USA.* **Robert W. Carpick** *is in the Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.*
e-mails: mathew.mate@hitachigst.com; carpick@seas.upenn.edu

1. Dowson, D. *Proc. Inst. Mech. Eng. J* **223,** 261–273 (2009).
2. Wandersman, E., Candelier, R., Debrégeas, G. & Prevost, A. *Phys. Rev. Lett.* **107,** 164301 (2011).
3. Romano, J. M., Hsiao, K., Niemeyer, G., Chitta, S. & Kuchenbecker, K. J. *IEEE Trans. Robotics* http://dx.doi.org/10.1109/tro.2011.2162271 (2011).
4. Mate, C. M. *Tribology on the Small Scale: A Bottom Up Approach to Friction, Lubrication, and Wear* 63–66 (Oxford Univ. Press, 2008).
5. Persson, B. N. J. *Sliding Friction: Physical Principles and Applications* (Springer, 1998).
6. Mate, C. M., McClelland, G. M., Erlandsson, R. & Chiang, S. *Phys. Rev. Lett.* **59,** 1942–1945 (1987).

CANCER

Sacrifice for survival

Cancer cells ignore oxygen availability, opting for less efficient, anaerobic ways of generating energy. The wisdom behind this choice seems to be in preventing the accumulation of reactive oxygen species, and so oxidative damage.

NANA-MARIA GRÜNING & MARKUS RALSER

When oxygen is plentiful, cells convert glucose to energy through the consecutive processes of glycolysis and oxidative respiration. However, cancer cells exhibit what is known as the Warburg effect: even in the presence of oxygen, they prefer the much less efficient process of glucose fermentation for energy production[1]. This seems counter-intuitive because rapid cell proliferation, which is required for tumour growth, has high energetic demands[2,3]. In a paper published in *Science*, Anastasiou and colleagues[4] provide evidence that cancer cells undergo this metabolic shift to clear reactive oxygen species (ROS) and so prevent oxidative damage. Thus, reconfiguration of the central carbon metabolism to counteract oxidative stress seems to be a major prerequisite for cancer progression.

Textbooks offer two possible explanations for the decline in respiratory activity during cancer development. First, with the increase in nucleotide and macromolecule biosynthesis, there is a shortage of carbon equivalents for oxidative respiration. Second, a higher speed of glycolysis makes anaerobic metabolism more efficient, with more lactate being generated from pyruvate, the end product of glycolysis; this allows cancer cells to feed each other by shuffling lactate[2,3].

Neither hypothesis fully explains the metabolic reconfiguration in cancer cells. For one,

© 2011 Macmillan Publishers Limited. All rights reserved

oxidative respiration occurs downstream of glycolysis, and so does not compete with glycolysis for carbon equivalents and would not interfere with a high glycolytic flux. Moreover, unlike respiring cells, which shuffle pyruvate from the cytoplasm into the mitochondria — the organelles within which oxidative respiration occurs — cancer cells actively excrete the lactate they generate from pyruvate. This contradicts the proposal that cancer cells shut down respiration to save carbon equivalents for biosynthesis. Finally, even some non-cancerous cells that do not make use of lactate (including yeast, T cells and induced pluripotent stem cells) undergo a Warburg-like metabolic restructuring during rapid growth.

Anastasiou and colleagues' results[4] bring the redox balance centre stage to explain this metabolic reconfiguration. They show that the glycolytic enzyme pyruvate kinase — a main regulator of the Warburg effect — facilitates tumour growth by preventing accumulation of ROS, and so avoiding oxidative damage.

In all living cells, ROS leak from the chain of reactions that constitute oxidative respiration, or are generated as by-products of both fatty-acid metabolism and biosynthetic redox reactions. Under normal physiological conditions this is not a problem, because ROS levels are kept low and in equilibrium with reducing molecules. In fact, a certain amount of ROS is necessary for normal physiology. But if the normal redox balance is disrupted, or ROS accumulate, oxidation and disturbed biochemical reactions damage macromolecules, ultimately leading to cell death. Therefore, cancer cells rely on a complex anti-oxidative machinery that can dynamically supply reducing equivalents and clear ROS when required[3].

Pyruvate kinase is a regulator of cellular anti-oxidative metabolism. Of the four human isoforms of this enzyme, PKM2 plays a crucial part in cancer metabolism. Like other metabolic enzymes, PKM2 levels increase in tumours[5]. However, this protein has a unique regulatory role in that its decreased catalytic activity is associated with tumour progression and the development of the Warburg effect[6,7].

When pyruvate kinase activity is low — as in cancer cells or in respiring yeast — its substrate, phosphoenol pyruvate, accumulates[8,9]. This inhibits the glycolytic enzyme triose phosphate isomerase and leads to activation of a pathway alternative to glycolysis — the pentose phosphate pathway[9]. Increased activity of this pathway protects against ROS in at least two ways. First, it provides NADPH, a reducing factor that is required for the activity of antioxidant enzymes and for the recycling of the anti-oxidant peptide glutathione. NADPH also compensates for the redox imbalance caused by increased nucleotide and fatty-acid synthesis[3]. Second, the pentose phosphate pathway regulates gene expression in favour of adaptation to oxidative stress[10].

Anastasiou and co-workers[4] establish that activation of the pentose phosphate pathway and its anti-oxidative activity are essential for cancer-cell growth (Fig. 1). They report that, in lung cancer cells, oxidation of PKM2 on the cysteine amino-acid residue 358 (Cys358) keeps its activity low. This increases both the concentration of glucose-6-phosphate — the metabolite that connects glycolysis to the oxidative, $NADP^+$-reducing branch of the pentose phosphate pathway — and flux through the pentose phosphate pathway.

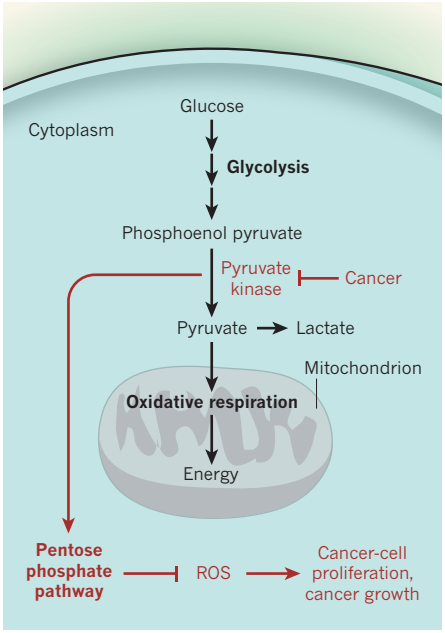


Figure 1 | Restructuring cellular metabolism. Glucose is converted to pyruvate by the cytoplasmic process of glycolysis, generating energy. When oxygen is present, pyruvate enters mitochondria, where it generates more energy through the process of oxidative respiration. But, in proliferating cells — and under anaerobic conditions — pyruvate is converted to lactate. In cancer and respiring yeast, reduced activity of pyruvate kinase, the enzyme that catalyses the final step of glycolysis, mediates redox balance by activating the pentose phosphate pathway[9]. Anastasiou *et al.*[4] show that activation of this pathway is crucial for cancer cells, and facilitates tumour growth by limiting ROS accumulation and, therefore, oxidative stress.

The authors interfered with the pyruvate-kinase-triggered activation of the pentose phosphate pathway by increasing PKM2 activity in the presence of oxidants. To do this, they mutated the enzyme's Cys358 to a serine residue or used small-molecule activators. This treatment had remarkable effects on cancer-cell growth. Accumulation of ROS caused oxidative damage and slowed the proliferation of cancer cells both in tissue culture and in tumours grafted into immunocompromised mice.

These data suggest that inducing the Warburg effect promotes cancer growth by activating the pentose phosphate pathway, maintaining the balance of redox equivalents, providing NADPH and activating antioxidant defence systems. The findings have notable implications for understanding the energetic balance during cancer development: blocking pyruvate kinase to redirect the metabolic flux is energetically costly under conditions of low respiratory activity because it diminishes the step that is responsible for the net yield of the cellular energy molecule ATP by glycolysis. This indicates that maintenance of the redox balance is more limiting for tumour growth than are energy levels or biosynthetic metabolism.

Could this metabolic reconfiguration be exploited for therapeutic purposes? Potentially, yes. But targeting a fundamental redox-balancing process must be cancer-cell specific, otherwise it would heavily damage other metabolically active cell types, including liver cells, immune cells and neurons. Yet, PKM2, triose phosphate isomerase, the pentose phosphate pathway and its associated metabolites are not cancer-cell specific. Nevertheless, a promising strategy might be to induce ROS overload in cancer cells, thereby making them vulnerable to oxidative damage by neutralizing the protective effects of the Warburg effect. To develop such strategies it will be essential to pursue comprehensive quantitative and qualitative investigations to understand all the ROS-producing biochemical reactions in the cancer cell. ■

Nana-Maria Grüning *and* **Markus Ralser** *are at the Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany. M.R. is also in the Department of Biochemistry and Cambridge Systems Biology Centre, University of Cambridge, Cambridge, CB2 1GA, UK. e-mails: gruening@molgen.mpg.de; mr559@cam.ac.uk*

1. Warburg, O. *Science* **123,** 309–314 (1956).
2. Hsu, P. P. & Sabatini, D. M. *Cell* **134,** 703–707 (2008).
3. Cairns, R. A., Harris, I. S. & Mak, T. W. *Nature Rev. Cancer* **11,** 85–95 (2011).
4. Anastasiou, D. *et al. Science* http://dx.doi.org/10.1126/science.1211485 (2011).
5. Bluemlein, K. *et al. Oncotarget* **2,** 393–400 (2011).
6. Hitosugi, T. *et al. Sci. Signal.* **2,** ra73 (2009).
7. Christofk, H. R. *et al. Nature* **452,** 230–233 (2008).
8. Vander Heiden, M. G. *et al. Science* **329,** 1492–1499 (2010).
9. Grüning, N.-M. *et al. Cell Metab.* **14,** 415–427 (2011).
10. Krüger, A. *et al. Antioxid. Redox Signal.* **15,** 311–324 (2011).

CORRECTION

In the News & Views article 'Ageing: Generations of longevity' by Susan E. Mango (*Nature* **479,** 302–303; 2011), it was stated that transient exposure of rats to a high-sugar/low-protein diet leads to glucose intolerance. This should have read "transient exposure of rats to a high-sugar/high-fat diet leads to glucose intolerance".

© 2011 Macmillan Publishers Limited. All rights reserved

REVIEW

doi:10.1038/nature10610

Assembly of hybrid photonic architectures from nanophotonic constituents

Oliver Benson[1]

The assembly of hybrid nanophotonic devices from different fundamental photonic entities—such as single molecules, nanocrystals, semiconductor quantum dots, nanowires and metal nanoparticles—can yield functionalities that exceed those of the individual subunits. Combining these photonic elements requires nanometre–scale fabrication precision and potentially involves a material diversity that is incompatible with standard nanotechnological processes. Although merging these different systems on a single hybrid platform is at present challenging, it promises improved performance and novel devices. Particularly rapid progress is seen in the combination of plasmonic–dielectric constituents with quantum emitters that can be assembled on demand into fundamental model systems for future optical elements.

The advancement of modern technology is still driven considerably by miniaturization. Following the path taken in electronics of reducing devices to their ultimately fundamental forms, for instance single-electron transistors[1–3], now magnetic and optical devices have also been scaled down, creating the increasingly active research fields of spintronics[4] and nanophotonics[5].

A fundamental nanophotonic device, involving single elementary excitations such as photons and spins, offers various advantages over a macroscopic object: its state can be controlled, initialized and read out with a precision at the fundamental quantum mechanical limit; its operational fidelity based on the interconversion of single excitations can approach 100%; and its functionality may be based on coherent rather than incoherent dynamics, allowing for devices that maintain quantum coherence, which is a crucial requirement for applications in quantum information processing[6]. Thus, a fundamental device could outperform traditional devices in terms of operation speed, integrability and energy efficiency.

On the fundamental level, a device consists of various coupled and/or integrated entities, such as single emitters, optical microresonators, optical waveguides and photon-to-electron or photon-to-spin interfaces. If different materials systems are involved (for example a combination of organic and inorganic units), then the resulting device might be appropriately described as 'hybrid'. However, this term can also be used in a broader sense when new functionality is gained from a combination of different physical effects. Examples are the fusion of bioassays and micro- or nanoelectromechanical systems[7] and that of microfluidics and optics[8]. A merging of quantum optics and nanomechanics[9], photonics and atomic physics[10], or quantum optics and plasmonics[11] is likely to provide new fundamental insight as well as novel applications such as quantum coherent devices[12] or quantum-limited sensors[13]. For example, a quantum computer, by contrast with its classical counterpart, needs to maintain coherent superpositions of discrete quantum states, acting as so-called quantum bits. A similar challenge occurs in quantum sensors, where exploitation of entangled quantum states allows for improved precision over classical equivalents. Whereas long storage times of quantum states may be possible, for example in atomic systems, fast read-out and information transfer using integrated optics, that is, an atomic–photonic hybrid system, would be desirable. The quantum mechanical decoherence and energy transfer processes in these hybrid systems are, however, not fully understood. This broader class of hybrid system is the topic of this Review, and I place a particular emphasis on model systems combining metallic and dielectric nanophotonic constituents by scanning probe manipulation.

I will begin by highlighting some key nanophotonic functionalities and then briefly introduce several bottom-up approaches for assembling hybrid nanophotonic systems from their fundamental components. As scanning probe techniques at present offer the greatest versatility, the successful application of such techniques for the assembly of hybrid nanophotonic architectures will be covered in most detail. Advanced top-down approaches based on layered composites[14], organic–inorganic synthesis[15], molecular recognition[16] or soft-matter assembly[17] will not be reviewed.

Functionality on the nanoscale

The intrinsic functionality of a nanophotonic device relies on it providing a significant interaction between light and matter. The most fundamental system consists of a single quantum emitter interacting with a single mode of the electromagnetic field, for example inside a microcavity[18]. Strong and efficient confinement of the electromagnetic field greatly enhances the field strengths and, thus, the strength of the interaction. There are basically two ways to confine the light and enhance its field strength: resonant enhancement in dielectric structures where cavity quantum electrodynamics[19] (CQED) effects can be exploited (Box 1), or plasmonic enhancement using metals[20] (Box 2). In a hybrid device, we might even engineer a combination of the two. Examples of desired or desirable functionalities are as follows.

Light guiding and sorting

A key requirement for a nanophotonic platform is for its different subunits, for example light sources, modulators, filters or detectors, to be connected by optical waveguides. This is a well-developed technique in standard light-wave communication networks and integrated optics. However, a nanophotonic circuit requires tighter confinement and low-intensity operation down to the single-photon level[21]. For dielectric systems, a particularly appropriate structure for nanophotonic devices is a photonic crystal[22], in which a periodic modulation of the index of refraction of a dielectric can guide light tightly and control its dispersion. Figure 1b shows such a device (discussed further below): a photonic

[1]Nano-Optik, Humboldt-Universität zu Berlin, Newtonstrasse 15, 12489 Berlin, Germany.

©2011 Macmillan Publishers Limited. All rights reserved

BOX 1

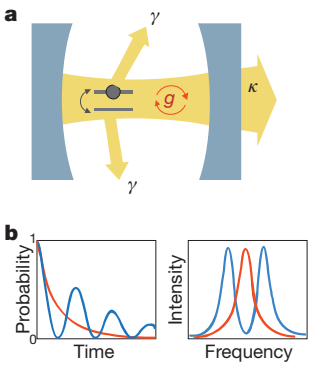
Cavity QED

A typical configuration in which to study CQED[19] is shown in the figure. A single quantum emitter, for example an atom, an ion, a single molecule or a quantum dot, is placed inside a microcavity. Two regimes of the time evolution of the system are distinguished.

In the strong-coupling regime, the dominant process is the coherent interaction between the emitter and a single resonant mode of the microcavity characterized by the coupling constant g. Energy is exchanged between the emitter and the cavity mode in an oscillatory way at the Rabi frequency, $2g$. The energy loss rate through either spontaneous emission from the excited emitter to a continuum of other modes (described by the rate γ) and the photon loss rate through the cavity mirrors (described by the rate κ) is much smaller than the energy exchange rate (γ and κ are both much less than g). The emitter and the field lose their separate identities and have to be described as one coupled system with an own energy-level structure[19]. The coupling constant g is proportional to $1/\sqrt{V}$, where V is the quantization (mode) volume. On the other hand, κ is inversely proportional to the cavity's quality factor Q. The ratio $Q/\sqrt{V}$ should therefore be maximized to reach the strong-coupling regime.

In the weak-coupling regime, the dynamics are incoherent. The coupling constant g is small relative to κ and γ. An excited emitter loses its energy in an exponential decay, but with a decay rate strongly modified relative to an emitter in free space. This is the Purcell effect[94]. The decay is enhanced when the emitter's transition frequency is in resonance with the cavity mode frequency, $\omega = c/\lambda$ (c, speed of light; λ, wavelength), or it may be strongly suppressed. For exact resonance and perfect alignment of the emitter's dipole moment with respect to the cavity field in a medium with refractive index n, the enhancement, or Purcell factor, is $P = (3/4\pi^2)(\lambda/n)^3 Q/V$.

CQED effects were first studied in atomic physics experiments[95], mostly using Fabry–Pérot cavities, but are now widely explored in solid-state systems as well[96].



Box 1 Figure | Basics of CQED. **a**, A typical CQED system consisting of a single quantum emitter (two-level system) coupled coherently, with coupling constant g, to one mode of an optical microcavity. Incoherent processes are spontaneous emission from the emitter to the continuum at rate γ, and photon loss through the cavity mirrors at rate κ. **b**, Time evolution of the excited-state probability and spectrum for the weak-coupling (red) and strong-coupling (blue) regimes.

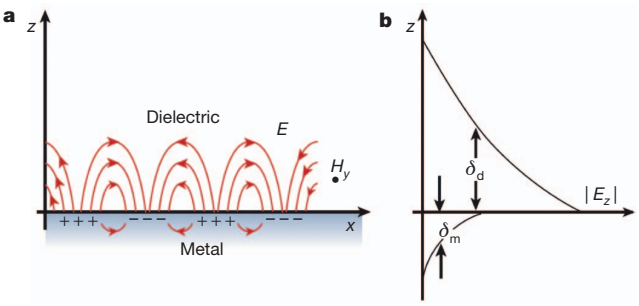
BOX 2

Plasmonic enhancement

Surface plasmon polaritons (SPPs) are electromagnetic excitations propagating at the interface between a dielectric and a conductor[20,24]. They arise from the coupling of electromagnetic fields to oscillations of the conductor's electron plasma. SPPs can be efficiently excited and guided along thin metal strips or wires, called SPP waveguides. Long-range propagation over distances of a few millimetres is possible at telecommunication wavelengths[24]. By comparison with their purely optical counterparts, the narrow spatial confinement to fractions of a free-space optical wavelength makes them ideal building blocks of future chip-based nanophotonic components and devices, although losses in the metal are a challenge. Passive optical elements such as bends, beam splitters and interferometers have been demonstrated[24].

Subwavelength metal particles support localized SPP resonances. With gold or silver, the resonances are in the visible spectral range and have been used for centuries to stain glass. Current applications concern biomarkers or sensors. In addition to a lightning rod effect, that is, the field enhancement near sharp geometrical features, resonances further increase the local electromagnetic field near the metallic surface. Thus, SPPs couple strongly to small particles placed in the near field. One feature is that fluorescence and scattering processes are enhanced by orders of magnitude. A prominent example with a pronounced impact on single-molecule detection is surface enhanced Raman scattering: in experiments on this, enhancement factors of 10^{14} have been observed[97]. Another feature is an increase in the efficiency of nonlinear optical effects for higher-harmonic generation[98] or for low-intensity photonic switches on the single-emitter level[41].

Modern nanofabrication technology makes it possible to fabricate plasmonic structures from the top down. More complex structures such as optical nanoantennae[32] and metallic photonic crystal structures[34], which further improve the performance of plasmonic nanophotonic elements[33], have also been realized.



Box 2 Figure | Basics of surface plasmon polaritons. **a**, An SPP as a collective excitation at a metal–dielectric interface[99]. The electromagnetic field (electric field, E, plotted in the z–x plane; magnetic field, H_y, sketched in the y direction) is drastically enhanced. **b**, The perpendicular field E_z decays exponentially with a characteristic length δ_d (of the order of the optical wavelength) in the dielectric and a characteristic length of δ_m (the skin depth) in the metal.

crystal waveguide guiding light within a thin dielectric film. Frequency-selective elements required, for example, for add–drop elements have also been realized in photonic crystal structures[23].

Whereas the confinement of light in a dielectric structure is ultimately constrained by the diffraction limit, plasmonic waveguides (Box 2) allow concentration of combined electronic and optical excitations (plasmon polaritons) to length scales far below the optical wavelength. Wave guiding is possible along plasmonic wires or along ridges and grooves in thin metal films[24]. A drawback is damping due to losses in the metal, but strategies to combine plasmonic waveguides with gain material in hybrid systems to compensate for losses are being explored[25]. The wavelength-selective transmission of plasmonic nanostructures can also be used for efficient light or photon sorting[26].

Enhanced emission and absorption

In the weak-coupling regime of CQED (Box 1), the Purcell effect can be used to funnel single or a few photons emitted from single emitters such as quantum dots, molecules or colour centres into specific cavity modes with enhanced rate and efficiencies of up to 98% (ref. 27). Efficient coupling to free space or to optical fibres is mandatory for improved

©2011 Macmillan Publishers Limited. All rights reserved

single-photon sources needed in secure quantum cryptography[28] or quantum intensity standards[29]. More-complex nanophotonic elements for these applications also implement funnelling of charge carriers for efficient electrical excitation of single quantum dots (Fig. 1a). The Purcell effect also significantly enhances the transition rate and, thus, the optical response of a device, which is relevant for fast optical modulation or interconnects.

It has been proposed by Chang *et al.*[30] that plasmonic effects will also lead to a strong enhancement of the interaction of single quantum emitters with localized excitations in plasmonic waveguides. A similar effect is exploited when enhancing the optical response of a plasmonic particle to the binding event of a large biomolecule, a technique that is used in plasmonic biosensors[31]. Also, the impedance mismatch between freely propagating light and confined light can be addressed using plasmonic structures. Plasmonic nanoantennas[32] can drastically enhance the emission or absorption rate, whereas periodic metallic nanostructures establish highly directional emission[33] or provide unusual linear properties, such as a negative index of refraction, unavailable in natural materials[34]. These features are of paramount importance for optical outputs or optical interconnects in future nanophotonic systems.

Nonlinear elements and switches

In the strong-coupling regime of CQED, spontaneous emission is negligible relative to the coherent and reversible energy exchange. This exchange is a perfect way to transfer electronic quantum states to photons. In terms of quantum information processing, this establishes an interface between stationary (electronic states) and moving (photons) quantum bits. Also, non-resonant interaction can be used to realize conditional phase shifts of photon states[19]. This is the main functionality required to build an optical quantum computer[35], which in principle could solve problems that no classical computer could[6].

The energy-level structure of a strongly coupled emitter–cavity system (the 'dressed states') is an array of manifolds of increasing energy, each of which corresponds to a discrete excitation of the system. A narrow-band, weak probe beam used to address a strongly coupled system can be on- or off-resonant with respect to the energy difference between two manifolds. In the first case it will be transmitted, but in the second case it will be reflected. As the energy of the manifold depends on the number of photons present in the cavity, the presence or absence of even a single photon can induce transmission (photon-induced tunnelling) or reflection (photon blockade) of the beam[36]. This peculiar effect allows the generation of transmitted light with a non-classical photon statistics[37], and, eventually, the realization of single-photon sources or transistors[36].

Waks and Vučković[38] have shown that a strong nonlinearity can be produced in a cavity–emitter system even if the system is not in the strong-coupling regime. In fact, a single emitter placed inside a cavity add–drop filter can switch on or off the transmission of a coupled waveguide mode as a result of interference effects. This 'dipole-induced transparency' may be used to build a quantum repeater[39], a device that helps to transfer quantum entanglement between two partners. Such a capability would establish fundamentally secure information exchange over arbitrarily large distances. Figure 1b shows a possible experimental configuration for dipole-induced transparency realized using a quantum dot and a photonic crystal structure[40].

In a similar way, a single emitter on a plasmonic waveguide can act as a mirror that can be saturated by only a single excitation. Such an element could be the key ingredient for a single-photon transistor[41] as sketched in Fig. 1d. It could provide the missing nonlinear element for photonic quantum gates, which have so far relied on interference and single-photon detection, for example in integrated photonic circuits[21] and telecommunication fibre systems[42].

A scaling up of fundamental devices will involve two or more elementary entities, for example single emitters, optical resonators or non-resonant optical nanostructures. Configurations may be arranged to act as optical switches, where the state of a single emitter changes the emission frequency of a nearby emitter by means of dipole–dipole interaction[43] (Fig. 1c).

Approaches to nanoassembling hybrid devices

To exploit fundamental interactions in nanophotonic–plasmonic hybrid structures, advanced nanoassembly techniques are required. Various strategies to fabricate fundamental hybrid devices out of different materials and subunits have been demonstrated. Figure 2 summarizes different approaches schematically.

Random assembly

The simplest idea is a random approach by which, for example, different nanoparticles are spin-coated or drop-cast on a substrate (Fig. 2a). By chance, the desired configuration may be found within the ensemble. This approach was used by, for example, Kinkhabwala *et al.*[44], who spin-coated a solution with molecules on a sample with 'bow-tie' plasmonic antennas. Sometimes they found a single molecule well within the

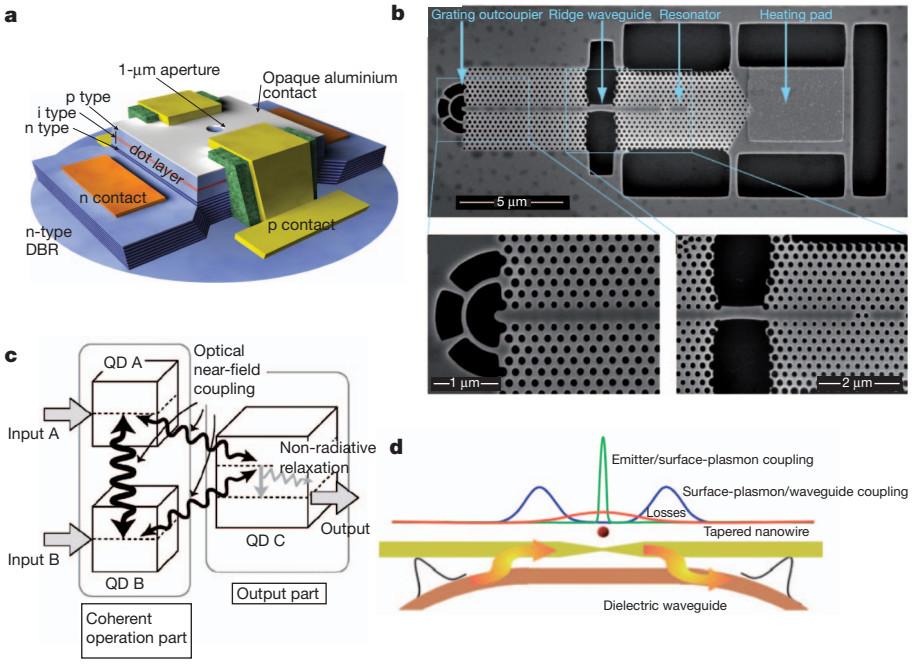


Figure 1 | Nanophotonic functional elements. **a**, Advanced, electrically driven single-photon source: a single semiconductor quantum dot embedded in a light-emitting-diode structure including a distributed Bragg reflector[88] (DBR). **b**, Scanning electron microscopy (SEM) image of an integrated photonic structure to test dipole-induced transparency. The structure includes a grating coupler (left; also shown magnified) and a quantum dot inside a photonic crystal resonator coupled to a photonic crystal waveguide[40] (also shown magnified). **c**, Proposed logic gate based on three near-field-coupled semiconductor quantum dots (QDs) with appropriately aligned energy levels[89]. Depending on the state of two near-field-coupled quantum dots (QD A and QD B, representing two bits) that can be controlled by optical input pulses, resonant energy transfer to a third quantum dot (QD C, serving as output) is possible. **d**, Proposal for a single-photon transistor[41]. A light pulse from a dielectric waveguide is transferred into a plasmon propagating along a tapered metal nanowire. In the tapered region, interaction with even a single emitter is so large that its quantum state state can control the transmission of the pulse.

©2011 Macmillan Publishers Limited. All rights reserved

antenna gap. Such an approach is very inefficient and cannot be scaled up to complex configurations involving three or more constituents[43].

Self-assembly

If there are interactions between the entities (submicrometre particles, large molecules and the like) that are strong enough to overcome thermal diffusion, then self-ordering may lead to deposition of the particles in ordered patterns (Fig. 2b). Van der Waals[45], electrostatic[14] or magnetic[46] interaction may be used to establish self-ordering. For example Min *et al.*[45] describe the formation of cobalt nanoparticle rings and cadmium nanorod superlattices under the influence of magnetic and electrostatic fields, respectively. The advantage of this technique is the control of patterns through modifications of the interactions and the possible fabrication of many configurations in parallel.

Directed self-assembly

Control of spatially ordered structures can be improved by growth or assembly on preprocessed or locally functionalized substrates (Fig. 2c). Etched defects or metal nanoparticles can act as seeds for self-assembled growth of single semiconductor quantum dots[47] or nanowires[48]. Mohan *et al.*[49] used electron beam lithography and etching to fabricate a regular pattern on a GaAs substrate. In this way, pyramidal quantum dots could be grown in a regular array. Also, patterning with chemical[50] markers allows deposition of specific molecules or nanoparticles on predefined sites. Such self-assembly approaches provide spatially well-separated structures that can be addressed individually. Apart from physical defects, local electric, magnetic or optical fields can also create favourable spots for the deposition of particles down to the single-atom level (Fig. 2d).

Select and post-process

This approach relies on the high precision of modern nanofabrication used, for example, for today's metal–oxide semiconductor technology. The approach that I here call 'select and post-process' starts with the fabrication of samples with a random distribution of nano-objects, for example optically active quantum emitters such as quantum dots or nanowires[51]. Subsequently, one object of interest is selected and its relative position on the sample is measured with high precision. Frequently, the substrate is patterned with an array of markers to facilitate this procedure. In a subsequent post-process step, electric contacts, microresonators or other elements are fabricated around or next to the selected object by optical or electron beam lithography (Fig. 2e). Using this technique, a fundamental nanophotonic system consisting of a single semiconductor quantum dot strongly coupled to an optical microresonator (a resonant defect mode in a photonic crystal structure) was realized[52]. As the select and post-process approach starts with a randomly distributed ensemble, it cannot be scaled up to assemble structures with more than one emitter. A possible solution, however, is post-processing of structures with site-controlled emitters[47].

Scanning probes for manual assembly

By comparison with the previously described approach, scanning probes offer the ultimate limit of nanomanipulation, allowing individual constituents to be assembled. First demonstrations with scanning tunnelling microscopes[53] (STMs) and atomic force microscopes[54] (AFMs) showed impressive results in regards to the positioning of single atoms[55] on surfaces.

STM and AFM scanning probes have developed into versatile tools for nanomanipulation and nanoassembly and are particularly useful in assembling hybrid structures. Not only large molecules[56], but also metallic nanoparticles[57] or nanorods[58], semiconductor nanoparticles[59], carbon nanotubes[60] and plastic particles[61], can be manipulated on a large variety of substrates. Modifications of 'dip-pen' nanolithography[62] make it possible to pick up and release nanoparticles in a controlled way[63] (Fig. 2f). Also, optical or scanning electron microscopes can be equipped with triple-axis micromanipulators for simultaneous imaging and manipulation.

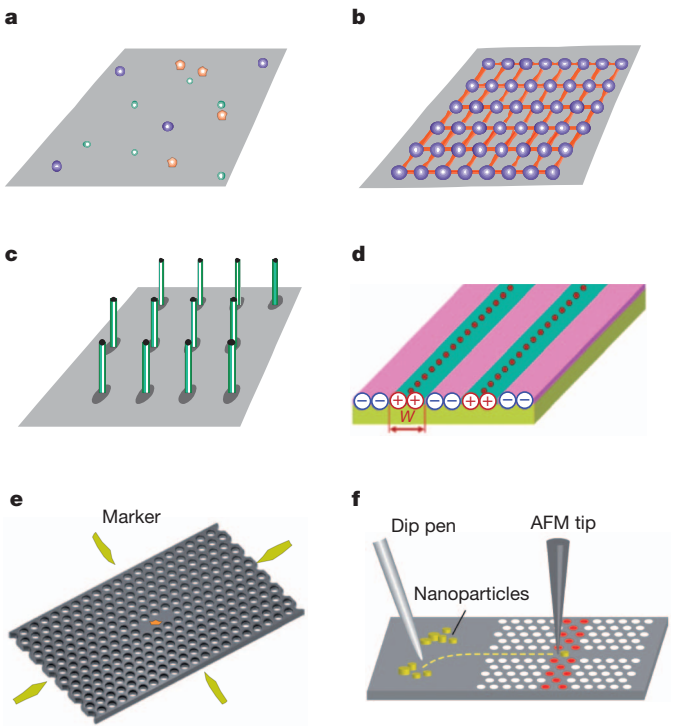


Figure 2 | Different approaches to assembling nanophotonic elements. **a**, Random approach: various particles are deposited randomly. By chance, two particles (here a yellow one and a green one) are close enough to form a coupled pair. **b**, Self-assembly: interparticle interactions, for example electrostatic fields (red lines), mediate a periodic configuration. **c**, Directed self-assembly (seeded growth): nanowires (green) are grown predominately on the prepatterned defects on a substrate. **d**, Directed self-assembly (field-guided assembly): negatively charged particles (red spheres) agglomerate along a positively charged wire pattern (green stripes of width *W*). **e**, Select and post-process: the position of a quantum dot is measured with respect to markers and subsequently a photonic crystal structure (periodic holes) are fabricated. **f**, Scanning probe assembly: particles deposited using a dip pen are precisely positioned with the help of an AFM tip.

Nanophotonic–plasmonic hybrid devices

Among the various approaches to assembling hybrid systems summarized above, scanning probe techniques provide the widest versatility, and so will be the focus of this section. Using such techniques, different configurations can be assembled with the same constituents, allowing a detailed investigation of the constituents' interactions. Furthermore, once assembled, the configurations can be encapsulated to be used as stable nanophotonic units in various applications, such as fibre-coupled single-photon sources[64] for quantum information processing or quantum light standards.

Whereas nanoassembly with scanning probes has been widely used in developing electronic[60], molecular[56] and biomolecular[56] devices, the assembly of nanophotonic elements has reached this point only recently and is still in its early stages. Progress has been hindered by the lack of photostable nanoscopic emitters that allow easy manipulation. At the level of very few or even single emitters under ambient conditions, protection against photobleaching by overgrowth or embedding in films is often required, preventing any additional positioning of the emitters on the substrate. Optically stable emitters appropriate for scanning probe assembly are metallic (plasmonic) nanoparticles[24], semiconductor nanowires[65] or colour centres in diamond nanocrystals[66,67]. Colour centres have particular potential as they represent single quantum emitters even at room temperature[68] (the most widely studied centre is the nitrogen–vacancy centre[66]).

©2011 Macmillan Publishers Limited. All rights reserved

To illustrate the power of scanning probes to assemble nanophotonic hybrid structures in a controlled way, I highlight two examples in more detail.

Plasmonically enhanced single-photon sources

Figure 3 shows a manually assembled, hybrid, single-photon source operating at room temperature[69]. To construct it, appropriate constituents (here metal nanoparticles and diamond nanocrystals) were first isolated and precharacterized. This was done with an AFM (in tapping mode, where only modest interaction with the particles occurs), to map the topology of a larger area, and with subsequent optical investigation, that is, measurements of fluorescence or photon correlations. This investigation provides information about the number of emitters in the particle[70]. This first step is similar to the select and post-process approach described above. However, in a second step active manipulation with an AFM tip (now in contact mode to exert significant force on the particles) was performed and complex configurations were fabricated in a step-by-step procedure (Fig. 3). This is sufficient for a well-defined coupling to plasmons, which requires an accuracy of some tens of an optical wavelength.

In this example, a single nanodiamond containing a single nitrogen–vacancy defect centre was coupled to a single gold particle or sandwiched between two gold particles forming a plasmonic nanoantenna. Plasmonic enhancement (see Box 2 and previous paragraph) leading to an increase in photon yield (number of photons emitted per unit of excitation power) of two orders of magnitude was achieved at room temperature. Even higher values should be possible[44] in a more advanced antenna structure[33], which should also allow directional emission. This excellent coupling of single emitters to plasmonic wires may be used in improved single-photon sources[30] to produce large nonlinearities for the realization of single-photon transistors[41] or to launch single plasmonic excitations into quantum plasmonic circuits.

Nanowire photonic elements

In another example (Fig. 4), various dielectric and plasmonic nanowires were manually assembled to form coupled lasers and waveguides[65]. The end facets of a small semiconductor nanowire can act as optical mirrors forming a Fabry–Pérot-type optical resonator. Optical excitation of the wire's photoluminescence can bring the system above the lasing threshold. In Fig. 4a, two such lasers, respectively made from GaN and ZnO wires, were coupled to a SnO_2 nanoribbon[71]. A triple-axis manipulator with a tungsten probe tip was used to arrange the configuration manually. The SnO_2 nanoribbon acts as an efficient single-mode waveguide to collect and guide the laser light both from the GaN wire and the ZnO wire. Remarkably, the coupling efficiency of laser light into the waveguide approaches 50%, and light can be guided with a loss of $\sim$10 dB cm^{-1}. Optimum coupling is achieved when wire and ribbon are in the 'staggered-bonded configuration', that is, in physical contact over distances of a few micrometres. Simultaneous guiding of light from two (or more) nanolasers establishes the possibility of performing all-nanowire nonlinear wave mixing within single nanocavities. Moreover, the ability to transfer coherent optical pulses over hundreds of micrometres could be applied to shuttle packets of electro-optical information in future computing and communications devices.

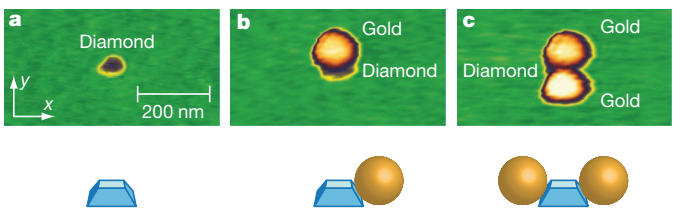


Figure 3 | Stepwise AFM nanoassembly of a plasmonic nanoantenna. a–c, A single nitrogen–vacancy defect centre in a diamond nanocrystal (**a**) is sandwiched between two gold nanoparticles (60 nm in diameter) (**b**, **c**). The sketches (bottom) illustrate the configuration mapped by false-colour AFM imaging (top).

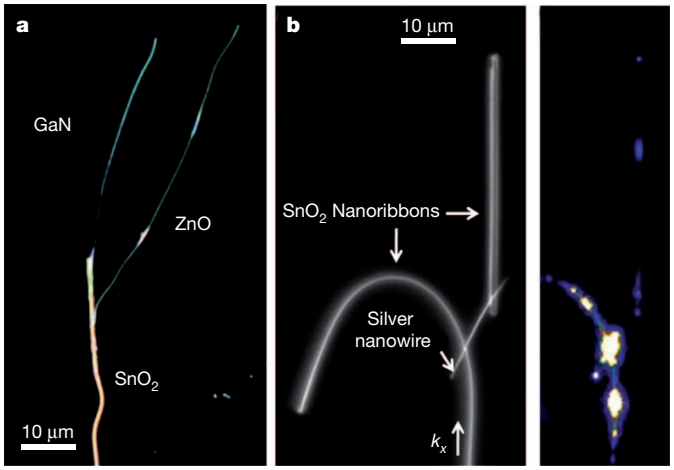


Figure 4 | Nanoassembled photonic–plasmonic wire structures. a, Dark-field image of two nanowire lasers (GaN and ZnO, each $\sim$130 nm in diameter) coupled to a SnO_2 nanoribbon waveguide[71] (with a rectangular cross-section of 240 nm $\times$ 260 nm). **b**, Dark-field (left) and optical (right) images of a hybrid optical–plasmonic routing device consisting of a silver nanowire bridging two SnO_2 nanoribbon waveguides[72]. Excitation light is incident from the bottom of the image (indicated by the arrow labelled k_x).

In Fig. 4b, it is shown how a silver plasmonic nanowire can be used to link two SnO_2 nanoribbon waveguides[72]. A transfer of light can be seen in the optical image. The system represents a hybrid optical–plasmonic routing device that combines the subwavelength confinement of light in a plasmonic structure with low-loss guiding in dielectric waveguides. Efficient integration and connection between plasmonic structures on a nanophotonic circuit can be realized in this way, which is a requirement for the single-photon transistor concept depicted in Fig. 1d. Connection of these structures to the 'outside world' with tapered optical fibres has also been demonstrated[72].

It must be noted that in the example in each of the two previous subsections, a hybrid approach is mandatory. In the first example there is no well-established fabrication technology for the material (advanced photonic structures in diamond), whereas in the second example the required properties of the material (low damping of the silver nanowire) can be established only by chemical synthesis, which is incompatible with lithographic fabrication. Nanomanipulation by means of micromanipulators[73,74] or advanced scanning probe techniques[75] complements the hybrid approach in an ideal way.

Future prospects

So far, several simple nanophotonic hybrid devices have been assembled with scanning probes or micromanipulators. Figure 5 shows some examples additional to those discussed in the previous section. Scaling up to produce more-complex nanophotonic elements in which a small but arbitrary number of emitters are controllably coupled to dielectric or plasmonic nanostructures is in principle straightforward.

A critical outstanding problem is to connect nanophotonic hybrid systems to macroscopic components. Tapered optical fibres[72] or plasmonic antenna structures[32] have proved to be excellent couplers to propagating light in optical waveguides and free beams, respectively, with efficiencies exceeding 90%. A strategy to relax the need for free-beam coupling is to incorporate sources and detectors on the nanophotonic chip itself. Examples of integrated emitters were provided in Figs 3–5 and on-chip detection of single surface plasmons was demonstrated recently[76].

It has to be mentioned that manual assembly with scanning probes has its own flaws. Emitters in solid-state matrices show an inhomogeneous broadening, with the result that in nanophotonic elements exploiting resonances, individual tuning of the emission frequency of each emitter is required. However, techniques for such individual tuning have been developed[77]. Also, emitters embedded in very small carrier particles, for example defect centres in diamond nanocrystals, often show spectral

©2011 Macmillan Publishers Limited. All rights reserved

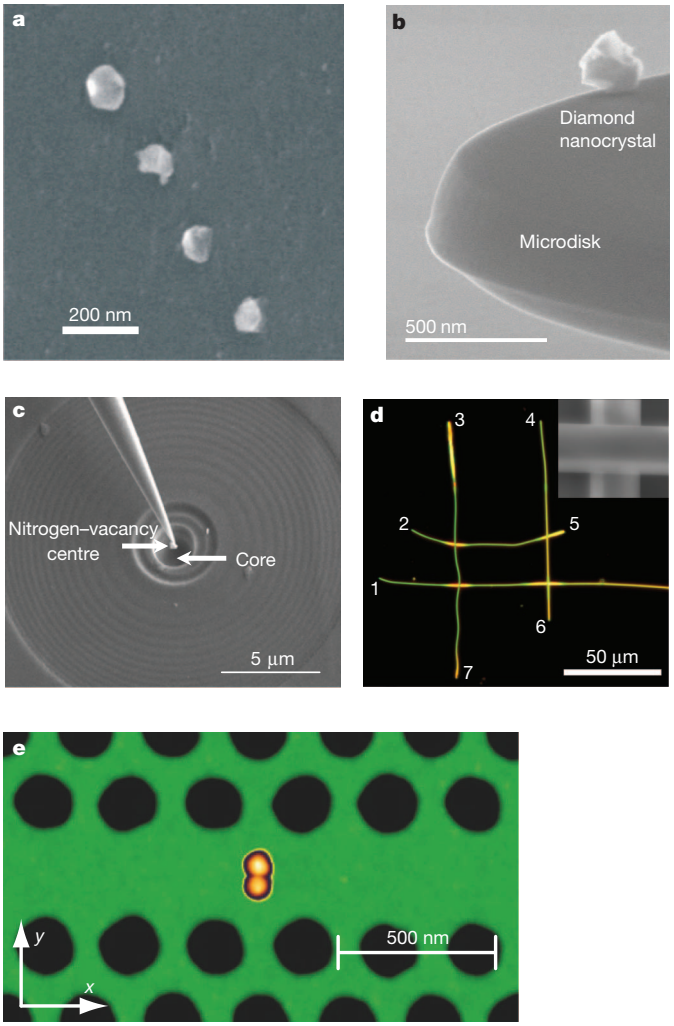


Figure 5 | Different systems assembled using scanning probes. **a**, SEM image of four nanodiamonds arranged in a linear chain[90]. **b**, SEM image of single nanodiamond, containing nitrogen–vacancy centres, placed on the rim of a SiO_2 microdisk resonator[91]. **c**, SEM image of a nanodiamond at the end of a nanomanipulator positioned on the end facet near the core region of an optical fibre[92]. With this set-up, fluorescence from the nitrogen–vacancy centre is coupled directly into the optical fibre[64]. **d**, Dark-field image SnO_2 nanoribbon grid[71]. Inset, scanning electron image of the junction at the lower right vertex. **e**, False-colour AFM image of two gold nanoparticles assembled to form a plasmonic nanoantenna on a photonic crystal heterostructure cavity etched into a thin Si_3N_4 membrane[93]

diffusion, blinking and linewidth broadening[78]. Furthermore, assembling a nanophotonic structure one constituent at a time with a precision of only a few nanometres is instructive for fundamental research and in proof-of-principle experiments, but it is not acceptable as a method of production of devices on a larger scale. However, the successful operation of the first nanophotonic hybrid model systems has motivated efforts to modify or improve established fabrication methods. For example, electron beam and optical lithography have been used to fabricate photonic crystal structures in materials in which this has not previously been possible (for example diamond[79] and GaP[80]), to develop plasmonic structures with reduced damping[81] and to realize integrated optical circuits for optical quantum technology[82]. Another approach is to combine different techniques with the goal of surpassing their individual limitations. One example of this is ion implantation through a perforated atomic force cantilever, which simultaneously provides nanometre spatial resolution and single-dopant operation[83].

The first key functionalities have been successfully demonstrated in hybrid photonic architectures assembled from fundamental nanophotonic constituents. Single photons can be generated, routed, detected and interconverted into other fundamental excitations such as single surface plasmons. Also, nonlinear interactions facilitating logical operations are feasible using CQED or plasmonic effects. There is great potential to establish additional functionality in a hybrid approach by combining optical, magnetic and electrical interactions. First steps towards nanomagnetometry[84], surface plasmon lasers[85], organic plasmon-emitting diodes[86] and single-molecule or single-photon transistors[87] have been taken.

Most of the systems realized so far represent model systems, but a remarkable level of complexity has already been reached. These manually assembled photonic architectures are an exciting test area in which to develop ideas for new functionalities in future nanophotonic hybrid elements.

1. Pothier, H. *et al.* Single electron pump fabricated with ultrasmall normal tunnel-junctions. *Physica B* **169,** 573–574 (1991).
2. Likharev, K. K. Single-electron devices and their applications. *Proc. IEEE* **87,** 606–632 (1999).
3. Klein, D. L., Roth, R., Lim, A. K. L., Alivisatos, A. P., & McEuen, P. L. A single-electron transistor made from a cadmium selenide nanocrystal. *Nature* **389,** 699–701 (1997).
4. Dietl, T., Awschalom, D. D., Kaminska, M. & Ohno, H. (eds) *Spintronics* (Semiconductors and Semimetals 82, Elsevier, 2008).
5. Prasad, P. N. *Nanophotonics* (Wiley, 2004).
6. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).
7. Huang, T. J. Recent developments in artificial molecular-machine-based active nanomaterials and nanosystems. *MRS Bull.* **33,** 226–231 (2008).
8. Psaltis, D., Quake, S. R. & Yang, C. Developing optofluidic technology through the fusion of microfluidics and optics. *Nature* **442,** 381–386 (2006).
9. Kippenberg, T. J. & Vahala, K. J. Cavity opto-mechanics. *Opt. Express* **15,** 17172–17205 (2007).
10. Folman, R., Krueger, P., Schmiedmayer, J., Denschlag, J. & Henkel, C. Microscopic atom optics: from wires to an atom chip. *Adv. At. Mol. Opt. Phys.* **48,** 263–356 (2002).
11. Akimov, A. V. *et al.* Generation of single optical plasmons in metallic nanowires coupled to quantum dots. *Nature* **450,** 402–406 (2007).
12. Cirac, J. I. & Zoller, P. A. A scalable quantum computer with ions in an array of microtraps. *Nature* **404,** 579–581 (2000).
13. Arcizet, O. *et al.* High-sensitivity optical monitoring of a micromechanical resonator with a quantum-limited optomechanical sensor. *Phys. Rev. Lett.* **97,** 133601 (2006).
14. Decher, G. & Schlenoff, J. B. (eds) *Multilayer Thin Films: Sequential Assembly of Nanocomposite Materials* (Wiley, 2003).
15. Yin, Y. & Alivisatos, P. Colloidal nanocrystal synthesis and the organic–inorganic interface. *Nature* **437,** 664–670 (2005).
16. Park, S. Y. *et al.* DNA-programmable nanoparticle crystallization. *Nature* **451,** 553–556 (2008).
17. You, C.-C., Verma, A. & Rotello, V. M. Engineering the nanoparticle-biomacromolecule interface. *Soft Matter* **2,** 190–204 (2006).
18. Vahala, K. J. Optical microcavities. *Nature* **424,** 839–846 (2003).
19. Berman, P. R. (ed.) *Cavity Quantum Electrodynamics* (Academic, 1993).
20. Maier, S. A. *Plasmonics: Fundamentals and Applications* (Springer, 2007).
21. Politi, A., Cryan, M. J., Rarity, J. G., Yu, S. & O'Brien, J. L. Silica-on-silicon waveguide quantum circuits. *Science* **320,** 646–649 (2008).
22. Joannopoulos, J. D. in *Photonic Crystals: Molding the Flow of Light* (Meade, R. D. & Winn, J. N.) (Princeton Univ. Press, 1995).
23. Song, B.-S., Noda, S. & Asano, T. Photonic devices based on in-plane hetero photonic crystals. *Science* **300,** 1537 (2003).
24. Bozhevolnyi, S. & García-Vidal, F. Focus on plasmonics. *N. J. Phys.* **10,** 105001 (2008).
25. Vasa, P. *et al.* Coherent exciton-surface-plasmon-polariton interaction in hybrid metal-semiconductor nanostructures. *Phys. Rev. Lett.* **101,** 116801 (2008).
26. Laux, E., Genet, C., Skauli, T. & Ebbesen, T. W. Plasmonic photon sorters for spectral and polarimetric imaging. *Nature Photon.* **2,** 161–164 (2008).
27. Englund, D., Faraon, A., Zhang, B., Yamamoto, Y. & Vuckovic, J. Generation and transfer of single photons on a photonic crystal chip. *Opt. Express* **15,** 5550–5558 (2007).
28. Gisin, N., Ribordy, G., Tittel, W. & Zbinden, H. Quantum cryptography. *Rev. Mod. Phys.* **74,** 145–195 (2002).
29. Grangier, P., Sanders, B. & Vuckovic, J. (eds) Focus on single photons on demand. *N. J. Phys.* **6** (2004).
30. Chang, D. E., Sørensen, A. S., Hemmer, P. A. & Lukin, M. D. Quantum optics with surface plasmons. *Phys. Rev. Lett.* **97,** 053002 (2006).
31. Anker, J. N. *et al.* Biosensing with plasmonic nanosensors. *Nature Mater.* **7,** 442–453 (2008).
32. Moerland, R. J., Taminiau, T. H., Novotny, L., van Hulst, N. F. & Kuipers, L. Reversible polarization control of single photon emission. *Nano Lett.* **8,** 606–610 (2008).
33. Curto, A. G. *et al.* Unidirectional emission of a quantum dot coupled to a nanoantenna. *Science* **329,** 930–933 (2010).
34. Smith, D. R., Pendry, J. B. & Wiltshire, M. C. K. Metamaterials and negative refractive index. *Science* **305,** 788–792 (2004).

©2011 Macmillan Publishers Limited. All rights reserved

35. Hofmann, H. F. & Takeuchi, S. Quantum phase gate for photonic qubits using only beam splitters and postselection. *Phys. Rev. A* **66,** 024308 (2002).
36. Imamoğlu, A., Schmidt, H., Woods, G. & Deutsch, M. Strongly interacting photons in a nonlinear cavity. *Phys. Rev. Lett.* **79,** 1467–1470 (1997).
37. Faraon, A. *et al.* Coherent generation of non-classical light on a chip via photon-induced tunnelling and blockade. *Nature Phys.* **4,** 859–863 (2008).
38. Waks, E. & Vučković, J. Dipole induced transparency in drop-filter cavity-waveguide systems. *Phys. Rev. Lett.* **96,** 153601 (2006).
39. Briegel, H. J., Dür, W., Cirac, I. & Zoller, P. Quantum repeaters: the role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81,** 5932–5935 (1998).
40. Faraon, A. *et al.* Dipole induced transparency in waveguide coupled photonic crystal cavities. *Opt. Express* **16,** 12154–12162 (2008).
41. Chang, D. E., Sørensen, A. S., Demler, E. A. & Lukin, M. D. A single-photon transistor using nanoscale surface plasmons. *Nature Phys.* **3,** 807–812 (2007).
This paper put forward a proposal for a single-photon transistor using plasmons.
42. Chen, J. *et al.* Demonstration of a quantum controlled-NOT gate in the telecommunications band. *Phys. Rev. Lett.* **100,** 133603 (2008).
43. Sangu, S., Kobayashi, K., Shojiguchi, A. & Ohtsu, M. Logic and functional operations using a near-field optically coupled quantum-dot system. *Phys. Rev. B* **69,** 115334 (2004).
44. Kinkhabwala, A. *et al.* Large single-molecule fluorescence enhancements produced by a bowtie nanoantenna. *Nature Photon.* **3,** 654–657 (2009).
This paper reported the enhancement of emission from single molecules coupled to a plasmonic antenna.
45. Min, Y., Akbulut, M., Kristiansen, K., Golan, Y. & Israelachvili, J. The role of interparticle and external forces in nanoparticle assembly. *Nature Mater.* **7,** 527–538 (2008).
46. Erb, R. M., Son, H. S., Samanta, B., Rotello, V. M. & Yellen, B. B. Magnetic assembly of colloidal superstructures with multipole symmetry. *Nature* **457,** 999–1002 (2009).
47. Sünner, T. *et al.* Scalable fabrication of optical resonators with embedded site-controlled quantum dots. *Opt. Lett.* **33,** 1759–1761 (2008).
48. Fan, H. J., Werner, P. & Zacharias, M. Semiconductor nanowires: from self-organization to patterned growth. *Small* **2,** 700–717 (2006).
49. Mohan, A. *et al.* Polarization-entangled photons produced with high-symmetry site-controlled quantum dots. *Nature Photon.* **4,** 302–306 (2010).
50. Jonas, U., del Campo, A., Krüger, C., Glasser, G. & Boos, D. Colloidal assemblies on patterned silane layers. *Proc. Natl Acad. Sci. USA* **99,** 5034–5039 (2002).
51. Park, H. G. *et al.* Wavelength-selective photonic-crystal waveguide coupled to a nanowire light source. *Nature Photon.* **2,** 622–626 (2008).
This paper reported a photonic structure using active semiconductor nanowires to light up photonic crystal waveguides.
52. Hennessy, K. *et al.* Quantum nature of a strongly coupled single quantum dot-cavity system. *Nature* **445,** 896–899 (2007).
53. Binnig, G., Rohrer, H., Gerber, Ch. & Weibel, E. Surface studies by scanning tunneling microscopy. *Phys. Rev. Lett.* **49,** 57–61 (1982).
54. Binnig, G., Quate, C. F. & Gerber, Ch. Atomic force microscope. *Phys. Rev. Lett.* **56,** 930–933 (1986).
55. Eigler, D. M. & Schweizer, E. K. Positioning single atoms with a scanning tunnelling microscope. *Nature* **344,** 524–526 (1990).
56. Kufer, S. K., Puchner, E. M., Gumpp, H., Liedl, T. & Gaub, H. E. Single-molecule cut-and-paste surface assembly. *Science* **319,** 594–596 (2008).
57. Tong, L., Zhu, T. & Liu, Z. Atomic force microscope manipulation of gold nanoparticles for controlled Raman enhancement. *Appl. Phys. Lett.* **92,** 023109 (2008).
58. Hsieh, S. *et al.* Imaging and manipulation of gold nanorods with an atomic force microscope. *J. Phys. Chem. B* **106,** 231–234 (2002).
59. Junno, T., Deppert, K., Montelius, L. & Samuelson, L. Controlled manipulation of nanoparticles with an atomic-force microscope. *Appl. Phys. Lett.* **66,** 3627–3629 (1995).
60. Thelander, C. & Samuelson, L. AFM manipulation of carbon nanotubes: realization of ultra-fine nanoelectrodes. *Nanotechnology* **13,** 108–113 (2002).
61. Harel, E., Meltzer, S. E., Requicha, A. A. G., Thompson, M. E. & Koel, B. E. Fabrication of polystyrene latex nanostructures by nanomanipulation and thermal processing. *Nano Lett.* **5,** 2624–2629 (2005).
62. Salaita, K., Wang, Y. & Mirkin, C. A. Applications of dip-pen nanolithography. *Nature Nanotechnol.* **2,** 145–155 (2007).
63. Wang, Y., Zhang, Y., Li, B., Lü, J. & Hu, J. Capturing and depositing one nanoobject at a time: single particle dip-pen nanolithography. *Appl. Phys. Lett.* **90,** 133102 (2007).
64. Schröder, T. *et al.* Fiber-integrated diamond-based single photon source. *Nano Lett.* **11,** 198–202 (2011).
65. Yan, R., Gargas, D. & Yang, P. Nanowire photonics. *Nature Photon.* **3,** 569–576 (2009).
66. Jelezko, F. & Wrachtrup, J. Single defect centres in diamond: a review. *Phys. Status Solidi A* **203,** 3207–3225 (2006).
67. Prawer, S. & Greentree, A. D. Applied physics - diamond for quantum computing. *Science* **320,** 1601–1602 (2008).
68. Kurtsiefer, C., Mayer, S., Zarda, P. & Weinfurter, H. Stable solid-state source of single photons. *Phys. Rev. Lett.* **85,** 290–293 (2000).
69. Schietinger, S., Barth, M., Aichele, T. & Benson, O. Plasmon-enhanced single photon emission from a nano-assembled metal-diamond hybrid structure at room-temperature. *Nano Lett.* **9,** 1694–1698 (2009).
This paper reported plasmon-enhanced emission of single photons from a single defect centre.
70. Sonnefraud, Y. *et al.* Diamond nanocrystals hosting single nitrogen-vacancy color centers sorted by photon-correlation near-field microscopy. *Opt. Lett.* **33,** 611–613 (2008).
71. Sirbuly, D. J. *et al.* Optical routing and sensing with nanowire assemblies. *Proc. Natl Acad. Sci. USA* **102,** 7800–7805 (2005).
72. Yan, R., Pausauskie, P., Huang, J. & Yang, P. Direct photonic-plasmonic coupling and routing in single nanowires. *Proc. Natl Acad. Sci. USA* **106,** 21045–21050 (2009).
This paper reported the excitation of plasmons by connecting silver nanowires to an active SnO_2 nanoribbon.
73. van der Sar, T. *et al.* Deterministic nanoassembly of a coupled quantum emitter–photonic crystal cavity system. *Appl. Phys. Lett.* **98,** 193103 (2011).
74. Englund, D. Deterministic coupling of a single nitrogen vacancy center to a photonic crystal cavity. *Nano Lett.* **10,** 3922–3926 (2010).
75. Wolters, J. *et al.* Enhancement of the zero phonon line emission from a single nitrogen vacancy center in a nanodiamond via coupling to a photonic crystal cavity. *Appl. Phys. Lett.* **97,** 141108 (2010).
76. Falk, A. L. *et al.* Near-field electrical detection of optical plasmons and single-plasmon sources. *Nature Phys.* **5,** 475–479 (2009).
This paper reported all-electrical plasmon detection using a field-effect transistor.
77. Hennessy, K. *et al.* Tuning photonic crystal nanocavity modes by wet chemical digital etching. *Appl. Phys. Lett.* **87,** 021108 (2005).
78. Shen, Y., Sweeney, T. M. & Wang, H. Zero-phonon linewidth of single nitrogen vacancy centers in diamond nanocrystals. *Phys. Rev. B* **77,** 033201 (2008).
79. Wang, C. F. *et al.* Fabrication and characterization of two-dimensional photonic crystal microcavities in nanocrystalline diamond. *Appl. Phys. Lett.* **91,** 201112 (2007).
80. Rivoire, K. *et al.* Lithographic positioning of fluorescent molecules on high-*Q* photonic crystal cavities. *Appl. Phys. Lett.* **95,** 123113 (2009).
81. Nagpal, P., Lindquist, N. C., Oh, S.-H. & Norris, D. J. Ultrasmooth patterned metals for plasmonics and metamaterials. *Science* **325,** 594–597 (2009).
82. O'Brien, J., Furusawa, A. & Vučković, J. Photonic quantum technologies. *Nature Photon.* **3,** 687–695 (2009).
This paper reviewed integrated quantum technologies based on optical chips and single photons.
83. Meijer, J. *et al.* Towards the implanting of ions and positioning of nanoparticles with nm spatial resolution. *Appl. Phys. A* **91,** 567–571 (2008).
84. Balasubramanian, G. *et al.* Nanoscale imaging magnetometry with diamond spins under ambient conditions. *Nature* **455,** 648–651 (2008).
85. Noginov, M. A. *et al.* Demonstration of a spaser-based nanolaser. *Nature* **460,** 1110–1112 (2009).
This paper reported the observation of spaser activity from a gold nanoparticle surrounded by a dye-doped silica shell.
86. Koller, D. M. *et al.* Organic plasmon-emitting diode. *Nature Photon.* **2,** 684–687 (2008).
This paper reported electrical excitation of plasmons through coupling to an organic light-emitting-diode structure.
87. Hwang, J. *et al.* A single-molecule optical transistor. *Nature* **460,** 76–80 (2009).
88. Bennett, A. J. *et al.* Microcavity single-photon-emitting diode. *Appl. Phys. Lett.* **86,** 181102 (2005).
89. Sangu, S., Kobayashi, K., Shojiguchi, A. & Ohtsu, M. Logic and functional operations using a near-field optically coupled quantum-dot system. *Phys. Rev. B* **69,** 115334 (2004).
90. van der Sar, T. *et al.* Nanopositioning of a diamond nanocrystal containing a single nitrogen-vacancy defect center. *Appl. Phys. Lett.* **94,** 173104 (2009).
91. Barclay, P. E., Santori, C., Fu, K.-M., Beausoleil, R. G. & Painter, O. Coherent interference effects in a nano-assembled optical cavity-QED system. *Opt. Express* **17,** 8081–8097 (2009).
92. Ampem-Lassen, E. *et al.* Nano-manipulation of diamond-based single photon sources. *Opt. Express* **17,** 11287–11294 (2009).
93. Barth, M. *et al.* Nanoassembled plasmonic-photonic hybrid cavity for tailored light-matter coupling. *Nano Lett.* **10,** 891–895 (2010).
This paper reported Fano-type resonances in a coupled system consisting of a photonic crystal cavity and gold nanoparticles.
94. Purcell, E. M. Spontaneous emission probabilities at radio frequencies. *Phys. Rev.* **69,** 681 (1946).
95. Raimond, J. M., Brune, M. & Haroche, S. Manipulating quantum entanglement with atoms and photons in a cavity. *Rev. Mod. Phys.* **73,** 565–582 (2001).
96. Yoshie, T. *et al.* Vacuum Rabi splitting with a single quantum dot in a photonic crystal nanocavity. *Nature* **432,** 200–203 (2004).
97. Kneipp, K. *et al.* Single molecule detection using surface-enhanced Raman scattering (SERS). *Phys. Rev. Lett.* **78,** 1667–1670 (1997).
98. Lippitz, M., van Dijk, M. A. & Orrit, M. Third-harmonic generation from single gold nanoparticles. *Nano Lett.* **5,** 799–802 (2005).
99. Barnes, W. L., Dereux, A. & Ebbesen, T. W. Surface plasmon subwavelength optics. *Nature* **424,** 824–830 (2003).

Acknowledgements I acknowledge financial supported by the Deutsche Forschungsgemeinschaft (BE2224/9, Sfb 951) and the Bundesministerium für Bildung und Forschung (KEPHOSI).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence should be addressed to the author (oliver.benson@physik.hu-berlin.de).

©2011 Macmillan Publishers Limited. All rights reserved

ARTICLE

doi:10.1038/nature10659

New gene functions in megakaryopoiesis and platelet formation

A full list of authors and their affiliations appears at the end of paper.

Platelets are the second most abundant cell type in blood and are essential for maintaining haemostasis. Their count and volume are tightly controlled within narrow physiological ranges, but there is only limited understanding of the molecular processes controlling both traits. Here we carried out a high-powered meta-analysis of genome-wide association studies (GWAS) in up to 66,867 individuals of European ancestry, followed by extensive biological and functional assessment. We identified 68 genomic loci reliably associated with platelet count and volume mapping to established and putative novel regulators of megakaryopoiesis and platelet formation. These genes show megakaryocyte-specific gene expression patterns and extensive network connectivity. Using gene silencing in *Danio rerio* and *Drosophila melanogaster*, we identified 11 of the genes as novel regulators of blood cell formation. Taken together, our findings advance understanding of novel gene functions controlling fate-determining events during megakaryopoiesis and platelet formation, providing a new example of successful translation of GWAS to function.

To discover novel genetic determinants of megakaryopoiesis and platelet formation, we performed meta-analyses of GWAS for mean platelet volume (MPV) and platelet count (PLT). Our analyses included 18,600 (13 studies, MPV) and 48,666 (23 studies, PLT) individuals of European descent, respectively, and up to ~2.5 million genotyped or imputed single nucleotide polymorphisms (SNPs)[1]. Briefly, we tested within each study (Supplementary Table 1) the associations of MPV and PLT with each SNP using an additive model; we then combined these study-specific test statistics in a fixed-effects meta-analysis. To reduce the risk of spurious associations, we applied common stringent quality control filters and the genomic control method[2] to the meta-analysis, which shows no evidence for residual inflation of summary statistics (Supplementary Fig. 1).

A total of 52 genomic loci reaching statistical significance at the genome-wide adjusted threshold of $P \leq 5 \times 10^{-8}$ were discovered in this stage 1 analysis; 55 additional loci reached suggestive association ($5 \times 10^{-8} < P \leq 5 \times 10^{-6}$). We tested one SNP per locus in a stage 2 analysis that included *in silico* and *de novo* replication data in up to 18,838 individuals from 12 additional studies, confirming 15 additional loci (Supplementary Table 2). One further independent locus (*TRIM58*) associated with PLT was identified through detection of secondary association signals. Overall, 68 independent genomic regions were associated with PLT and MPV with $P \leq 5 \times 10^{-8}$, of which 52 are new and 16 were described previously in Europeans[3–6] (Table 1). Of the 68 loci, 43 and 25 loci were associated significantly with PLT and MPV, respectively; 16 of them reached genome-wide significance with both traits (Supplementary Fig. 2). This partial overlap reflects the negative correlation of both traits (gender-adjusted $r = -0.49$, Fig. 1a) that results from the tight control of platelet mass (PLT × MPV)[7]. The association of some loci with both PLT and MPV may reflect this negative correlation between the two traits or independent pleiotropic effects of a locus on megakaryopoiesis and platelet formation. The different statistical power at the two traits and small effect sizes at many loci reduce our power to discriminate among loci controlling MPV and PLT through analysis of platelet mass. Their testing will require the collection and analysis of PLT and MPV in large independent homogeneous cohorts. Some loci, however, have a clear-cut effect. For instance, *BAK1* affects PLT specifically, compatible with its role in apoptosis and platelet lifespan.

We further tested the association of the 68 loci in 7,949 (MPV) and 8,295 (PLT) samples of south Asian and 14,697 (PLT) samples of Japanese[8] origin. We detected substantial overlap of association signals, with effect size and direction highly concordant with findings in Europeans (Supplementary Fig. 3 and Supplementary Table 3). In the south Asian sample, 15 of the 68 (22.1%) loci were significant after adjustment for multiple testing ($P \leq 7 \times 10^{-4}$). In the Japanese sample, 13 of 55 (23.6%) PLT loci showed significance. Moreover, 73 of 84 (87%, South Asians) and 45 of 55 (82%, Japanese) SNPs showed associations with effect estimates directionally consistent with Europeans. Such concordance is highly unlikely to be due to chance ($P = 2.3 \times 10^{-12}$ and $P = 2.1 \times 10^{-6}$), and provides independent validation of the locus discovery in Europeans.

The 68 loci cumulatively explain 4.8% of the phenotypic variance in PLT and 9.9% in MPV, accounting respectively for average increases of $2.57 \times 10^9 l^{-1}$ PLT and 0.10 fl MPV per copy of allele. These levels of explained variance are in accordance with other GWAS of complex quantitative traits[9]. Our results indicate that many other common variants of similar or lower effect size, rare variants as well as structural variants may also contribute to the variation of both platelet traits. We used the method of ref. 10 to estimate the number of additional PLT- and MPV-associated loci having effect sizes comparable to those observed in our analysis. The method (with caveats discussed in the Supplementary Information) predicted that 137 and 81 such loci exist for PLT and MPV respectively, accounting for 9.7% and 18.3% of the total phenotypic variance.

Gene-prioritization strategies

Evidence from recent, highly powered meta-analyses suggests that the association peaks are enriched for genes controlling key underlying biological pathways[11,12]. In our case, a large proportion of the association signals (46 out of 68) had the most significant SNP in stage 1 ('sentinel SNP') mapping to within a gene-coding region, including several key regulators of haemostasis (*ITGA2B*, *F2R*, *GP1BA*), megakaryopoiesis (*THPO*, *MEF2C*) and platelet lifespan (*BAK1*). Through an unbiased analysis of our GWAS results, we estimated that PLT-associated SNPs are significantly more likely to map to gene regions than expected by chance ($P < 0.05$, Supplementary Fig. 4), suggesting that we may prioritize the search of additional yet unknown genes

©2011 Macmillan Publishers Limited. All rights reserved

Table 1 | Summary of loci associated with platelet count and mean platelet volume in Europeans

GWAS locus	Trait	Sentinel SNP	Chr (build 36)	Position (build 36)	Cytoband	Locus	Effect/other allele*	*n*	Effect (s.e.)†	*P* value‡	Het. *P* value	Rep.¶	Refs#
1	MPV	rs17396340	1	10,208,762	1p36.22	*KIF1B*	A/G	21,612	0.008 (0.002)	2.83×10^{-8}	0.83	–	–
2	PLT	rs2336384	1	11,968,649	1p36.22	*MFN2*	G/T	57,366	2.172 (0.382)	1.25×10^{-8}	0.31	–	–
3	MPV	rs10914144	1	170,216,372	1q24.3	*DNM3*	C/T	18,589	0.014 (0.001)	1.11×10^{-24}	0.46	Yes	6
	PLT	rs10914144	1	170,216,372			T/C	54,978	3.417 (0.487)	2.22×10^{-12}	0.79	Yes	
4	MPV	rs1172130	1	203,511,575	1q32.1	*TMCC2*	G/A	21,141	0.011 (0.001)	3.82×10^{-27}	0.17	Yes	6
	PLT	rs1668871	1	203,503,759			C/T	58,108	2.804 (0.368)	2.59×10^{-14}	0.45	–	
5	PLT	rs7550918	1	245,742,181	1q44	LOC148824	T/C	54,171	3.133 (0.471)	2.91×10^{-11}	0.85	–	–
6	PLT	rs3811444	1	246,106,073	1q44	*TRIM58*§	C/T	27,955	3.346 (0.574)	5.60×10^{-9}	0.66	–	–
7	PLT	rs1260326	2	27,584,443	2p23	*GCKR*	T/C	54,396	2.334 (0.381)	9.12×10^{-10}	0.11	Yes	–
8	MPV	rs649729	2	31,317,888	2p23.1	*EHD3*	T/A	20,850	0.008 (0.001)	1.17×10^{-12}	0.61	Yes	6
	PLT	rs625132	2	31,335,803			G/A	45,217	4.236 (0.568)	9.15×10^{-14}	0.98	–	
9	PLT	rs17030845	2	43,541,382	2p21	*THADA*	C/T	65,738	3.577 (0.556)	1.27×10^{-10}	0.40	–	–
10	MPV	rs4305276	2	241,143,685	2q37.3	*ANKMY1*	G/C	20,618	0.008 (0.001)	1.71×10^{-11}	0.71	–	–
11	PLT	rs7616006	3	12,242,647	3p25	*SYN2*	A/G	58,564	1.997 (0.366)	4.86×10^{-8}	0.20	–	–
12	PLT	rs7641175	3	18,286,415	3p23	*SATB1*	A/G	58,366	2.757 (0.416)	3.37×10^{-11}	0.34	–	–
13	MPV	rs1354034	3	56,824,788	3p14.3	*ARHGEF3*	T/C	18,286	0.023 (0.001)	3.31×10^{-69}	0.00	Yes	4,6
	PLT	rs1354034	3	56,824,788			C/T	49,135	6.848 (0.442)	2.86×10^{-54}	0.50	Yes	
14	PLT	rs3792366	3	124,322,565	3q21.1	*PDIA5*	G/A	58,335	2.153 (0.365)	3.60×10^{-9}	0.07	–	–
15	MPV	rs10512627	3	125,822,911	3q21.1	*KALRN*	C/G	21,108	0.006 (0.001)	5.10×10^{-10}	0.41	–	–
16	MPV	rs11734132	4	6,942,419	4p16.1	*KIAA0232*	G/C	17,444	0.011 (0.002)	1.11×10^{-11}	0.20	–	–
17	PLT	rs7694379	4	88,405,532	4q22.1	*HSD17B13*	A/G	56,430	2.129 (0.37)	8.70×10^{-9}	0.44	Yes	–
18	MPV	rs2227831	5	76,059,249	5q13.3	*F2R*	G/A	21,654	0.021 (0.003)	9.65×10^{-16}	0.11	–	–
	PLT	rs17568628	5	76,082,694			T/C	44,759	6.074 (0.993)	9.61×10^{-10}	0.77	–	
19	PLT	rs700585	5	88,187,872	5q14.3	*MEF2C*	C/T	55,469	2.703 (0.442)	9.86×10^{-10}	0.06	–	–
	MPV	rs4521516	5	88,135,706			G/C	28,157	0.008 (0.001)	1.89×10^{-9}	0.39	–	
20	PLT	rs2070729	5	131,847,819	5q31.1	*IRF1*	A/C	56,469	2.394 (0.371)	1.13×10^{-10}	0.73	–	–
21	MPV	rs10076782	5	158,537,540	5q33.3	*RNF145*	A/G	18,025	0.007 (0.001)	4.48×10^{-8}	0.52	–	–
22	PLT	rs441460	6	25,656,266	6p22.2	*LRRC16*	G/A	58,064	3.08 (0.359)	8.70×10^{-18}	0.61	–	–
23	PLT	rs3819299	6	31,430,345	6p21.33	*HLA-B*	G/T	48,687	5.048 (0.824)	8.80×10^{-10}	0.90	–	–
24	PLT	rs399604	6	33,082,991	6p21.32	*HLA-DOA*	C/T	57,674	2.346 (0.365)	1.30×10^{-10}	0.23	–	–
25	PLT	rs210134	6	33,648,186	6p21.31	*BAK1*	G/A	58,554	4.957 (0.396)	7.11×10^{-36}	0.67	Yes	6,8
26	PLT	rs9399137	6	135,460,710	6q23.3	*HBS1L–MYB*	C/T	57,857	5.901 (0.41)	5.04×10^{-47}	0.74	Yes	8
27	MPV	rs342293	7	106,159,454	7q22.3	*FLJ36031–*	G/C	20,193	0.017 (0.001)	7.03×10^{-57}	0.19	Yes	5,6
	PLT	rs342275	7	106,146,451		*PIK3CG*	C/T	58,571	3.742 (0.363)	5.57×10^{-25}	0.17	–	
28	PLT	rs4731120	7	123,198,458	7q31.3	*WASL*	C/A	66,147	4.14 (0.592)	2.77×10^{-12}	0.46	–	–
29	PLT	rs6993770	8	106,650,703	8q23.1	*ZFPM2*	A/T	54,960	3.668 (0.437)	4.30×10^{-17}	0.14	–	–
30	PLT	rs6995402	8	145,077,548	8q24.3	*PLEC1*	C/T	57,593	2.304 (0.371)	5.09×10^{-10}	0.10	–	–
31	MPV	rs10813766	9	321,489	9p24.3	*DOCK8*	T/G	21,104	0.007 (0.001)	3.68×10^{-12}	0.45	–	–
32	PLT	rs409801	9	4,734,742	9p24.1	*AK3*	C/T	56,063	5.585 (0.378)	2.59×10^{-49}	0.47	–	6
33	PLT	rs13300663	9	4,804,947	9p24.1	*RCL1*	C/G	48,092	5.585 (0.483)	9.83×10^{-30}	0.64	Yes	8
34	PLT	rs3731211	9	21,976,846	9p21.3	*CDKN2A*	A/T	54,529	3.281 (0.438)	6.43×10^{-14}	0.86	Yes	–
35	PLT	rs11789898	9	135,915,483	9q34.2	*BRD3*	T/G	57,391	3.014 (0.476)	2.39×10^{-10}	0.70	–	–
36	MPV	rs7075195	10	64,720,664	10q21.2	*JMJD1C*	A/G	21,226	0.014 (0.001)	2.39×10^{-44}	0.89	Yes	6
	PLT	rs10761731	10	64,697,615			T/A	54,344	3.849 (0.378)	2.02×10^{-24}	0.56	Yes	
37	PLT	rs505404	11	233,267	11p15.5	*PSMD13–*	G/T	54,642	4.662 (0.453)	7.44×10^{-25}	0.86	–	6
	MPV	rs17655730	11	260,714		*NLRP6*	T/C	20,875	0.01 (0.001)	2.29×10^{-15}	0.29	–	
38	PLT	rs4246215	11	61,320,874	11q12.2	*FEN1*	T/G	56,299	2.451 (0.39)	3.31×10^{-10}	0.41	Yes	–
39	PLT	rs4938642	11	118,605,115	11q23.3	*CBL*	C/G	56,605	4.73 (0.727)	7.66×10^{-11}	0.98	Yes	–
40	MPV	rs1558324	12	6,159,479	12p13.31	*CD9–VWF*	A/G	20,387	0.01 (0.001)	1.55×10^{-21}	0.48	Yes	–
	PLT	rs7342306	12	6,161,353			G/A	55,636	2.532 (0.384)	4.29×10^{-11}	0.14	–	
41	MPV	rs2015599	12	29,326,746	12p11.22	*MLSTD1*	A/G	21,102	0.008 (0.001)	5.55×10^{-16}	0.76	–	–
42	MPV	rs10876550	12	52,998,574	12q13.13	*COPZ1–*	G/A	21,214	0.008 (0.001)	1.86×10^{-14}	0.38	Yes	–
						NFE2–CBX5							
43	MPV	rs2950390	12	55,341,557	12q13.3	*PTGES3–*	C/T	21,238	0.008 (0.001)	7.45×10^{-14}	0.75	–	–
	PLT	rs941207	12	55,309,550		*BAZ2A*	G/C	55,653	2.751 (0.431)	1.74×10^{-10}	0.33	–	
44	PLT	rs3184504	12	110,368,990	12q24.12	*SH2B3*	T/C	56,354	3.99 (0.374)	1.22×10^{-26}	0.07	–	6,8
45	PLT	rs17824620	12	111,585,376	12q24.13	*RPH3A–*	C/A	51,530	2.457 (0.428)	9.67×10^{-9}	0.26	–	–
						PTPN11							
46	MPV	rs7961894	12	120,849,965	12q24.31	*WDR66*	T/C	29,755	0.03 (0.001)	1.42×10^{-103}	0.48	Yes	4,6
	PLT	rs7961894	12	120,849,965			C/T	51,897	3.923 (0.609)	1.22×10^{-10}	0.05	Yes	
47	PLT	rs4148441	13	94,696,207	13q32	*ABCC4*	G/A	64,120	4.117 (0.6)	6.76×10^{-12}	0.38	Yes	–
48	MPV	rs7317038	13	113,060,898	13q34	*GRTP1*	C/T	27,646	0.006 (0.001)	8.27×10^{-12}	0.09	–	–
49	PLT	rs8022206	14	67,590,658	14q24.1	*RAD51L1*	G/A	52,251	3.197 (0.5)	1.55×10^{-10}	0.59	–	–
50	PLT	rs8006385	14	92,570,778	14q31	*ITPK1*	G/A	64,929	3.587 (0.558)	1.24×10^{-10}	0.28	–	–
51	PLT	rs7149242	14	100,229,168	14q32.2	*C14orf70–*	G/T	61,247	2.142 (0.385)	2.68×10^{-8}	0.07	–	–
						DLK1							
52	PLT	rs11628318	14	102,109,839	14q32.31	*RCOR1*	A/T	62,438	2.572 (0.405)	2.04×10^{-10}	0.84	–	–
53	PLT	rs2297067	14	102,636,537	14q32.32	*C14orf73*	T/C	41,687	3.538 (0.553)	1.58×10^{-10}	0.79	–	–
	MPV	rs944002	14	102,642,567			A/G	22,910	0.008 (0.001)	4.76×10^{-11}	0.66	Yes	
54	MPV	rs3000073	14	104,800,836	14q32.33	*BRF1*	G/A	21,229	0.007 (0.001)	3.27×10^{-11}	0.28	–	–
55	PLT	rs3809566	15	61,120,776	15q22.2	*TPM1*	G/A	57,113	2.443 (0.39)	3.65×10^{-10}	0.33	–	6
56	PLT	rs1719271	15	62,970,853	15q22.31	*ANKDD1A*	G/A	56,782	3.414 (0.502)	1.05×10^{-11}	0.00	–	–
57	PLT	rs6065	17	4,777,160	17pter-p12	*GP1BA*	T/C	64,987	4.191 (0.63)	2.92×10^{-11}	0.00	Yes	8
58	PLT	rs397969	17	19,744,838	17p11.1	*AKAP10*	C/T	60,944	2.131 (0.357)	2.32×10^{-9}	0.92	–	–
59	MPV	rs8076739	17	24,738,712	17q11.2	*TAOK1*	T/C	21,652	0.013 (0.001)	4.59×10^{-38}	0.12	–	4,6
	PLT	rs559972	17	24,838,621			T/C	53,460	3.264 (0.375)	3.30×10^{-18}	0.25		

©2011 Macmillan Publishers Limited. All rights reserved

Table 1 | Continued

GWAS locus	Trait	Sentinel SNP	Chr (build 36)	Position (build 36)	Cytoband	Locus	Effect/other allele*	*n*	Effect (s.e.)†	*P* value‡	Het. *P* value	Rep.¶	Refs#
60	PLT	rs10512472	17	30,908,916	17q12	*SNORD7–AP2B1*	C/T	58,692	3.636 (0.477)	2.40×10^{-14}	0.08	Yes	–
	MPV	rs16971217	17	30,968,167			C/G	21,089	0.009 (0.001)	3.77×10^{-12}	0.01		
61	PLT	rs708382	17	39,797,869	17q21.31	*FAM171A2–ITGA2B*	T/C	50,036	2.439 (0.431)	1.51×10^{-8}	0.46	–	–
62	PLT	rs11082304	18	18,974,970	18q11.2	*CABLES1*	G/T	58,215	2.48 (0.378)	5.27×10^{-11}	0.73	Yes	–
63	MPV	rs12969657	18	65,687,475	18q22.2	*CD226*	T/C	19,285	0.007 (0.001)	3.36×10^{-11}	0.38	Yes	6
64	MPV	rs8109288	19	16,046,558	19p13.12	*TPM4*	A/G	13,964	0.029 (0.004)	1.15×10^{-11}	0.14	–	3
	PLT	rs8109288	19	16,046,558			G/A	29,014	11.945 (1.892)	2.75×10^{-10}	0.04	–	
65	PLT	rs17356664	19	50,432,610	19q13.32	*EXOC3L2*	C/T	55,487	2.599 (0.415)	3.60×10^{-10}	0.07	–	–
66	MPV	rs13042885	20	1,872,706	20p13	*SIRPA*	C/T	21,186	0.008 (0.001)	5.56×10^{-14}	0.56	–	6
67	MPV	rs4812048	20	57,021,165	20q13.32	*CTSZ–TUBB1*	C/T	20,811	0.008 (0.001)	1.30×10^{-9}	0.06	–	–
68	PLT	rs1034566	22	18,364,276	22q11.21	*ARVCF*	T/C	61,469	2.128 (0.384)	3.06×10^{-8}	0.43	–	–
69	PLT	rs6141	3	185,572,959	3q27	*THPO*‖	T/C	39,366	2.467 (0.456)	6.18×10^{-8}	0.59	Yes	24, 8

Results are provided for the 68 loci and 84 sentinel SNPs reaching genome-wide significant ($P \leq 5 \times 10^{-8}$) association with PLT or MPV. Results for stages 1 and 2 of the analysis in Europeans are provided in Supplementary Table 2. MPV, mean platelet volume; PLT, platelet count.

* Alleles are indexed to the forward strand of NCBI build 36.

† Effect sizes in ln(fl) for MPV and $10^9 l^{-1}$ for PLT.

‡ All *P* values are based on the inverse-variance weighted meta-analysis model (fixed effects).

§ *TRIM58* identifies the only secondary signal identified in this study, derived from a genome-wide secondary signal discovery effort carried out by conditioning the discovery GWAS on all SNPs reaching significance in the stage 1 meta-analysis. The effects (s.e.) and *P* values reported are obtained in the secondary analysis. The corresponding values in the stage 1 analysis are effect (s.e.) = 2.721 (0.542) and $P = 4.06 \times 10^{-7}$. Further details of this analysis are given in the Supplementary Information.

‖ *THPO* narrowly misses the level required for nominal significance ($P < 5 \times 10^{-8}$) in Europeans, but shows genome-wide significance in Japanese.

¶ Rep. indicates replication of European stage 1+2 results in non-Europeans (Supplementary Table 3): yes, if association *P* value is at least in one non-European population <0.0007 (to account for multiple testing of 68 loci).

#Relevant references are indicated.

controlling these processes in the associated regions. To define a univocal rule to study the enrichment of functional relationships in associated genes, we made the choice to focus on a set of 54 'core' genes selected as either containing the sentinel SNP or mapping to within 10 kb from an intergenic sentinel SNP (Table 2). This selection strategy is designed to obtain unbiased hypotheses producing interpretable biological inference for genes near the association signals, but has reduced sensitivity for genes that map further from the sentinel SNP. For instance *VWF*, a key regulator of haemostasis, maps to 55 kb from the sentinel SNP (Supplementary Fig. 3 and Supplementary Table 4) and is therefore not considered as a core gene. We further note that this selection strategy does not imply knowledge of the location of causative

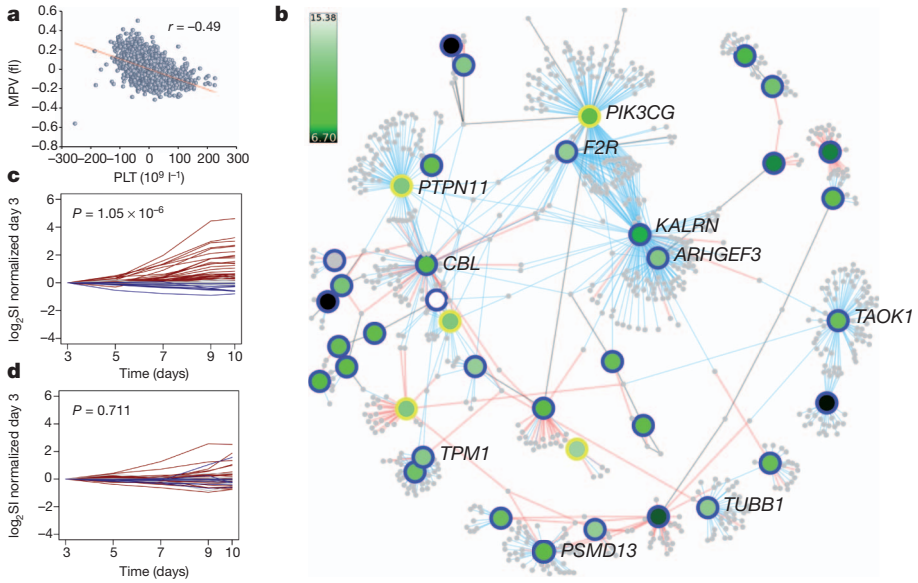


Figure 1 | Protein–protein interaction network and gene transcription patterns. a, Negative correlation between PLT and MPV in a UK sample. The gender-adjusted correlation coefficient *r* and trend line are shown. **b**, Protein–protein interaction network of platelet loci. For the nodes, genes are represented by round symbols, where node colour reflects gene transcript level in megakaryocytes on a continuous scale from low (dark green) to high (white). Grey-coloured round symbols identify first-order interactors identified in Reactome and IntAct. Core genes not connected to the main network are omitted. The 34 core genes are identified by a blue perimeter. Yellow perimeters identify five additional genes (*VWF*, *PTPN11*, *PIK3CG*, *NFE2* and *MYB*) with known roles in haemostasis and megakaryopoiesis and mapping to within the association signals at distances greater than 10 kb from the sentinel SNPs. These genes, which do not conform to the rule for inclusion into the core gene list, are not considered in further analyses presented in Fig. 2c, d and Supplementary Fig. 5 and are shown here for illustration purposes only. Network edges were obtained from the Reactome (blue) and IntAct-like (red) databases and through manual literature curation (black). The network including the 34 core genes alone contains 633 nodes and 827 edges; after inclusion of the 5 additional genes, the network (shown here) includes 785 nodes and 1,085 edges. The full network, containing gene expression levels and other annotation features, is available in Cytoscape[25] format for download (Supplementary Data 1). **c**, **d**, Time course experiments of gene expression in megakaryocytes and erythroblasts. Expression of core genes in $\log_2$ transformed signal intensities ($\log_2$ SI) during differentiation of the haematopoietic stem cells into megakaryocytes (**c**) or erythroblasts (**d**), segregated by their trends of statistically significant increasing (red), decreasing (blue) or unchanged (grey) gene expression. The corresponding gene list for the three classes is given in Supplementary Data 1.

©2011 Macmillan Publishers Limited. All rights reserved

Table 2 | Summary of functional evidence for core genes

Sentinel SNP (trait)	Core gene (distance in kb)*	Phenotype†
rs17396340 (MPV)	*KIF1B* kinesin family member 1B (0)	Variant annotation: sentinel SNP in $r^2 = 1$ with eQTL for *K1F1B*
rs2336384 (PLT)	*MFN2* mitofusin 2 (0)	Variant annotation: sentinel SNP is eQTL for *MFN2*
rs10914144 (MPV, PLT)	*DNM3* dynamin 3 (0)	*shibire* (*DNM-like*): overproliferation of plasmatocytes in *Drosophila* (this study)
rs1172130, rs1668871 (MPV, PLT)	*TMCC2* transmembrane and coiled-coil domain family 2 (2,481; 0)	Variant annotation: sentinel SNP in $r^2 = 0.928$ with eQTL for *RIPK5*
rs1260326 (PLT)	*GCKR* glucokinase (hexokinase 4) regulator (0)	*Xab1* (non-core gene): pronounced increase in plasmatocyte and crystal cell counts in *Drosophila* (this study)
rs649729, rs625132 (MPV, PLT)	*EHD3* EH-domain-containing 3 (0)	*ehd3* morpholino-injected embryos had no haematopoietic phenotype in *D. rerio* (this study)
rs17030845 (PLT)	*THADA* thyroid adenoma associated (0)	*Zfp36l2* (non-core gene): decreased platelet cell number in mouse
rs1354034 (MPV, PLT)	*ARHGEF3* Rho guanine nucleotide exchange factor (GEF) 3 (0)	*arhgef3*: profound effect on thrombopoiesis and erythropoiesis in *D. rerio* (this study)
rs2227831, rs17568628 (MPV, PLT)	*F2R* coagulation factor II (thrombin) receptor (0; 15,343)	*Par1* (*F2R*): thrombin activation of platelets attenuated in mouse
rs700585, rs4521516 (PLT, MPV)	*MEF2C* myocyte enhancer factor 2C (0)	*Mef2*: severely impaired megakaryopoiesis with reduced platelet count and increased platelet volume in mouse
rs2070729 (PLT)	*IRF1* interferon regulatory factor 1 (0)	*Irf1*: decreased number of NK lymphocytes in mouse
rs10076782 (MPV)	*RNF145* ring finger protein 145 (0)	*rnf145*: ablation of thrombopoiesis and erythropoiesis in *D. rerio* (this study)
rs210134 (PLT)	*BAK1* BCL2-antagonist/killer 1 (116)	*Bak1*: genetic ablation of *Bcl-xl* in mouse leads to thrombocytopenia by reducing platelet lifespan and this is corrected by ablation of *Bak1*
rs6993770 (PLT)	*ZFPM2* zinc finger protein, multitype 2 (0)	*Zfpm2*: peripheral haemorrhage in mouse; *ush* (*ZFPM2*): reduction in plasmatocytes and crystal cells in *Drosophila* (this study)
rs6995402 (PLT)	*PLEC1* plectin (0)	*Plec1*: impaired leukocyte recruitment to wounds in mouse
rs10813766 (MPV)	*DOCK8* dedicator of cytokinesis 8 (0)	*Dock8*: decrease in number of B cells and T cells in mouse; autosomal recessive hyper-IgE recurrent infection syndrome (OMIM: 243700)
rs409801 (PLT)	*AK3* adenylate kinase 3 (2,699)	*ak3*: ablation of thrombopoiesis and erythropoiesis in *D. rerio* (this study)
rs7075195, rs10761731 (MPV, PLT)	*JMJD1C* jumonji domain containing 1C (0)	*jmjd1c*: ablation of thrombopoiesis and erythropoiesis in *D. rerio* (this study)
rs505404, rs17655730 (PLT, MPV)	*PSMD13* proteasome (prosome, macropain) 26S subunit, non-ATPase, 13 (0); *NLRP6* NLR family, pyrin domain containing 6 (7,856)	*rpn9* (*PSMD13*): reduction in plasmatocyte numbers in *Drosophila* (this study)
rs4246215 (PLT)	*FEN1* flap structure-specific endonuclease 1 (0)	Variant annotation: sentinel SNP is eQTL for *CPSF7; Fads2* (non-core gene): abnormal platelet physiology and decreased platelet aggregation in mouse
rs4938642 (PLT)	*CBL* Cas-Br-M (murine) ecotropic retroviral transforming sequence (0)	Acute myeloid leukaemia (OMIM: 165360); *Cbl*: increased platelet numbers, increased thymic CD3 and CD4 expression on T cells in mouse; haematopoietic stem/progenitor cells showed enhanced sensitivity to cytokines in *Cbl*-null mice
rs10876550 (MPV)	*COPZ1* coatomer protein complex, subunit zeta 1 (6,604)	Variant annotation: sentinel SNP is eQTL for *GPR84*; *Copz1*: iron deficiency in mouse; *Nfe2* (non-core): thrombocytopenia in mouse; *Znf385a* (non-core): abnormal platelet morphology in mouse; *Su(var)205* (non-core *CBX5*): reduction in plasmatocyte number and overproliferation of crystal cells in *Drosophila* (this study)
rs3184504 (PLT)	*SH2B3* SH2B adaptor protein 3 (0)	*Lnk* (*SH2B3*): increased megakaryopoiesis and platelet count in mouse; increased white blood cell counts and decreased platelet count in mouse; *rpl6* (non-core): reduced plasmatocyte and crystal cell number in *Drosophila* (this study)
rs4148441 (PLT)	*ABCC4* ATP-binding cassette, sub-family C (CFTR/MRP), member 4 (0)	*ABCC4* is an active constituent of mediator-storing granules in human platelets
rs7317038 (MPV)	*GRTP1* growth hormone regulated TBC protein 1 (0)	Variant annotation: sentinel SNP is eQTL for *GRTP1*; sentinel SNP in $r^2 = 0.93$ with eQTL for *RASA3*
rs3000073 (MPV)	*BRF1* BRF1 homologue, subunit of RNA polymerase III transcription initiation factor IIIB (*S. cerevisiae*) (0)	*brf*: reduction in plasmatocyte cell number in *Drosophila* (this study)
rs3809566 (PLT)	*TPM1* tropomyosin 1 (alpha) (1,115)	Variant annotation: sentinel SNP in $r^2 = 1$ with eQTL for *TPM1*; *tpma* (*TPM1*): total abrogation of thrombopoiesis, but normal erythropoiesis in *D. rerio* (this study)
rs6065 (PLT)	*GP1BA* glycoprotein Ib (platelet), alpha polypeptide (0)	Bernard–Soulier syndrome (OMIM: 231200), benign Mediterranean macrothrombocytopenia (OMIM: 153670), pseudo-von Willebrand disease (OMIM: 177820); *Gp1ba*: giant platelets, a low platelet count and increased bleeding in mouse; $Gp1ba^{+/-}$ mice show complete inhibition of arterial thrombus formation and intermediate platelet numbers
rs708382 (PLT)	*FAM171A2* family with sequence similarity 171, member A2 (1,108); *ITGA2B* integrin, alpha 2b (7,207)	Glanzmann thrombasthenia (OMIM: 607759); decreased platelet count, abnormal platelet morphology and decreased platelet aggregation in mouse; *itga2b*: severely reduced thrombocyte function in *D. rerio*
rs12969657 (MPV)	*CD226* CD226 molecule (0)	Variant annotation: sentinel SNP is eQTL for *CD226*; leukocyte adhesion deficiency (OMIM: 116920); CD226 mediates adhesion of megakaryocytic cells to endothelial cells and inhibition of this diminishes megakaryocytic cell maturation; *Dnam1* (*CD226*): cytotoxic T cells and NK cells less able to lyse tumours in mouse
rs13042885 (MPV)	*SIRPA* signal-regulatory protein alpha (4,163)	*Sirpa*: mild thrombocytopenia in mouse, decreased proportion of single positive T cells, enhanced peritoneal macrophage phagocytosis
rs4812048 (MPV)	*CTSZ* cathepsin Z (5,468); *TUBB1* tubulin, beta 1 (6,539)	Autosomal dominant macrothrombocytopenia (OMIM: 613112); *Tubb1*: thrombocytopenia resulting from a defect in generating proplatelets in mouse; prolonged bleeding time and attenuated response of platelets to thrombin in mouse; *sun* (non-core ATP5E): reduction of crystal cell numbers in *Drosophila* (this study)

©2011 Macmillan Publishers Limited. All rights reserved

Table 2 | Continued

Sentinel SNP (trait)	Core gene (distance in kb)*	Phenotype†
rs1034566 (PLT)	*ARVCF* armadillo repeat gene deleted in velocardiofacial syndrome (0)	Variant annotation: sentinel SNP in $r^2 = 1$ with eQTL for *UFD1L*
rs6141 (PLT)	*THPO* thrombopoietin (0)	Essential thrombocythemia (OMIM: 187950); *Thpo*: decrease in platelet number and increase in platelet volume in mouse

Information is given only for genes with a haematopoietic phenotype. A more extensive annotation of genes within associated intervals is presented in Supplementary Table 4. Information on variants associated with gene expression is presented in Supplementary Table 6. No evidence for a haematopoietic effect was associated with the following core genes: rs3811444 (PLT) (*TRIM58* tripartite motif-containing 58 (0)); rs4305276 (MPV) (*ANKMY1* ankyrin repeat and *MYND* domain containing 1 (0)); rs3792366 (PLT) (*PDIA5* protein disulphide isomerase family A, member 5 (0)); rs10512627 (MPV) (*KALRN* kalirin, RhoGEF kinase (0)); rs11734132 (MPV) (*KIAA0232* (5,628)); rs441460 (PLT) (*LRRC16A* leucine-rich-repeat containing 16A (0)); rs13300663 (PLT) (*RCL1* RNA terminal phosphate cyclase-like 1 (0)); rs3731211 (PLT) (*CDKN2A* cyclin-dependent kinase inhibitor 2A (0)); rs2950390, rs941207 (MPV, PLT) (*PTGES3* prostaglandin E synthase 3 (cytosolic) (1,871); *BAZ2A* bromodomain adjacent to zinc finger domain, 2A (0)); rs7961894 (MPV, PLT) (*WDR66* WD repeat domain 66 (0)); rs8022206 (PLT) (*RAD51L1* RAD51-like 1 (*S. cerevisiae*) (0)); rs8006385 (PLT) (*ITPK1* inositol-tetrakisphosphate 1-kinase (0)); rs2297067, rs944002 (PLT, MPV) (*C14orf73* exocyst complex component 3-like 4 (0)); rs8076739, rs559972 (MPV, PLT) (*TAOK1* TAO kinase 1 (3,357, 0)); rs1697127 (MPV) (*AP2B1* adaptor-related protein complex 2, beta 1 subunit (0)); rs11082304 (PLT) (*CABLES1* Cdk5 and Abl enzyme substrate 1 (0)); rs8109288 (MPV, PLT) (*TPM4* tropomyosin 4 (0)), rs17356664 (PLT) (*EXOC3L2* exocyst complex component 3-like 2 (3,301)); rs2015599 (MPV) (*FAR2* fatty acyl CoA reductase 2 (0)); rs397969 (PLT) (*AKAP10* A kinase (PRKA) anchor protein 10 (4,506)); rs11789898 (PLT) (BRD3 bromodomain containing 3 (0)).

* Core genes are defined as either containing a sentinel SNP or as mapping at less than 10 kb from an intergenic SNP. Distance from nearest gene is calculated as the absolute distance between SNP and transcription start site of the gene or 3′ end of last exon.

† Phenotypes are defined from exhaustive search of the OMIM (Online Mendelian Inheritance in Man) database, published *in vitro* studies for humans and knockout and knockdown experiments for model organisms for both core and non-core genes. r^2 values are calculated from the HapMap phase 2 CEU panel. *Drosophila* indicates *Drosophila melanogaster*.

variants, which is currently incomplete. A detailed SNP survey showed that at 15 loci the sentinel SNPs either encoded, or were in high linkage disequilibrium (LD, $r^2 \geq 0.8$) with, a non-synonymous variant (Supplementary Table 5); another 11 either matched or were in high linkage disequilibrium with SNPs associated with expression levels of core genes (or cis-eQTLs, Supplementary Table 6), indicating that other loci may exert their effect through regulation of gene expression[13]. The validation of suggestive causative effects, as well as the identification of more complex interactions involving other genomic loci (trans eQTLs), will require a more comprehensive discovery in appropriately powered genomic data sets.

As a first effort to characterize biological connectivity among the core genes, we applied canonical pathway analyses (see http://www.ingenuity.com), detecting a highly significant over-representation of core genes in relevant biological functions such as haematological disease, cancer and cell cycle (Supplementary Table 7). Encouraged by these results, we extended this effort to construct a comprehensive network of protein–protein interactions incorporating the core genes. This effort integrated information from public databases (principally Reactome and IntAct) with careful manual revision of published evidence and high-throughput gene expression data. The resulting network, which includes 633 nodes and 827 edges, showed extensive connectivity between the proteins encoded by the core genes with an established functional role in megakaryopoiesis and platelet formation and those encoded by genes hitherto unknown to be implicated in these processes (Fig. 1b).

Transcriptional patterns of core genes

We next considered whether this connectivity was also reflected in the regulation of core gene transcription, and whether expression patterns were unique to megakaryocytes. Despite high levels of correlation in gene expression between different blood cell types (median = 0.8; median absolute deviation = 0.1)[14], we found that core genes tend to have significantly greater expression in megakaryocytes than in the other blood cells ($P = 7.5 \times 10^{-5}$, Supplementary Fig. 5a). This observation is compatible with the notion that ultimate steps in blood cell lineage specification are accompanied, or driven, by the emergence of increasing numbers of lineage-specific transcripts. To explore this assumption, we used genome-wide expression arrays to determine changes in global transcript levels during *in vitro* differentiation of umbilical-cord blood-derived haematopoietic stem cells to precursors of blood cells. We considered five different time points and two cell types, erythroblasts (the precursors of red blood cells) and megakaryocytes. Notwithstanding high levels of correlation of gene expression between erythroblasts and megakaryocytes[14], core gene transcripts showed a significant increase over time in megakaryocytes ($P = 1.5 \times 10^{-6}$) but not in erythroblasts ($P = 0.77$, Fig. 1c, d; see also Supplementary Fig. 5b). Taken together, these patterns of core gene expression are consistent with a different regulation of their transcription in megakaryocytes versus erythroblasts, and with their centrality in megakaryopoiesis and platelet formation. This hypothesis is also consistent with the observation that only 5 of the 68 sentinel SNPs exert a significant effect on erythrocyte parameters (*HBS1L-MYB*, *RCL1*, *SH2B3*, *TRIM58* and *TMCC2*, Supplementary Table 8).

Gene silencing in model organisms

To assess whether core genes are indeed implicated in haematopoiesis, we interrogated the function of 15 genes using gene silencing in *D. rerio* and *D. melanogaster*, and supported empirical data with published evidence on knockout models in *M. musculus* (Table 2 and Supplementary Table 4). In *D. rerio*, we applied morpholino constructs to silence the expression of six genes (Fig. 2 and Supplementary Fig. 6) selected to have >50% homology with the human counterpart and no previous evidence of involvement in haematopoiesis. Silencing of four genes in *D. rerio* (*arhgef3*, *ak3*, *rnf145*, *jmjd1c*) resulted in the ablation of both primitive erythropoiesis and thrombocyte formation. Silencing of *tpma*, the orthologue of *TPM1* that is transcribed in megakaryocytes but not in other blood cells, abolished the formation of thrombocytes but not of erythrocytes. Silencing of *ehd3* did not yield a haematopoietic phenotype. We also screened *D. melanogaster* RNA interference (RNAi) knockdown lines for quantitative alterations in the two most prevalent classes of blood elements: plasmatocytes and crystal cells. The repertoire of blood cells in *D. melanogaster*, consisting of about 95% plasmatocytes and 5% crystal cells, is less varied than in vertebrates. Transcription factors and signalling pathways regulating haematopoiesis have, however, been conserved throughout evolution[15], making the RNAi knockdown studies a relevant first step towards a better understanding of the putative role of these GWAS genes in haematopoiesis. Four core-gene *D. melanogaster* lines (*shibire* (*DNM*), *ush* (*ZFPM2*), *rpn9* (*PSMD13*), *Brf* (*BRF1*)), as well as five others (*sun* (*ATP5E*), CG3704 (*XAB1*), *Su(var)205* (*CBX5*), *dve* (*SATB1*) and *RpL6* (*RPL6*)), displayed highly reproducible differences in the numbers of these two cell types (Table 2 and Supplementary Table 4). Despite widespread differences between mammalian and insect haematopoietic lineages[16], our findings from *D. melanogaster* provide new and supporting examples of functional conservation in the control of blood cell formation in invertebrates and vertebrates[17–19].

New gene and functional discoveries

The data from studies in *D. rerio* by us and in *M. musculus* by others (see Supplementary Table 4) provided proof-of-concept evidence that our prioritization strategy is appropriate for selecting novel genes controlling thrombopoiesis and megakaryopoiesis, respectively. More detailed insights and additional implicated genes will be revealed through the systematic silencing of all genes in the associated regions. For instance, RNAi knockdown of *dve* in *D. melanogaster* reduces plasmatocyte numbers and increases the number of crystal cells, thus providing supporting evidence that its non-core gene

©2011 Macmillan Publishers Limited. All rights reserved

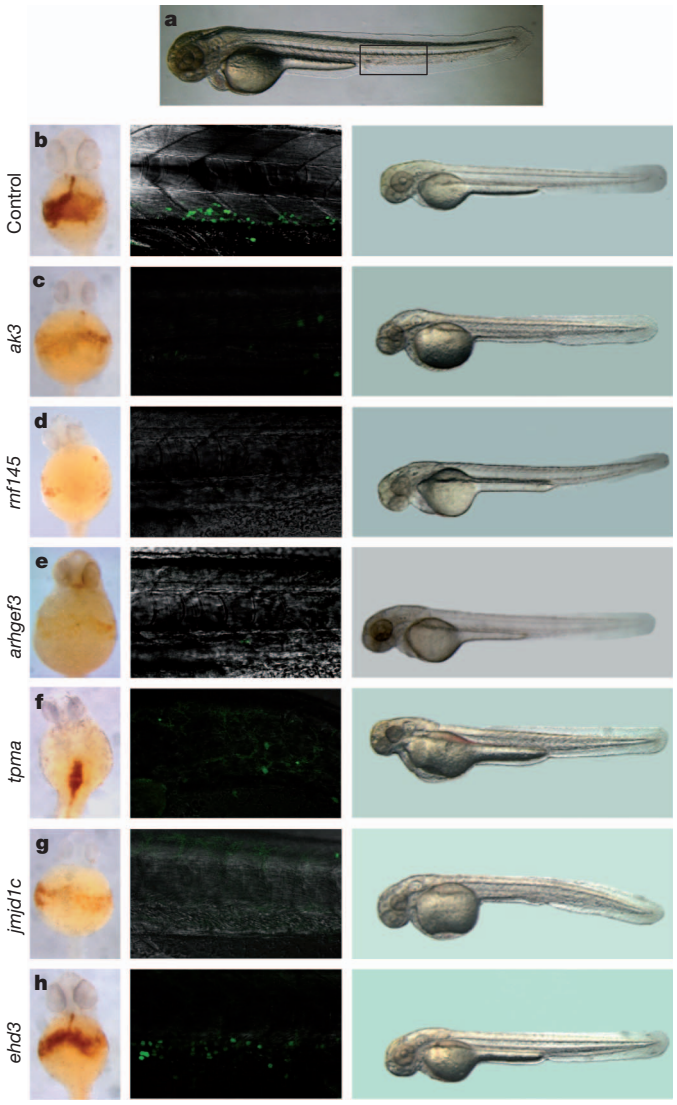


Figure 2 | Functional assessment of novel loci in *D. rerio*. Gene-specific morpholinos were injected into wild-type and *Tg(cd41:EGFP)* embryos at the one cell stage (Supplementary Fig. 6) to assess alterations in erythropoiesis and thrombopoiesis. **a**, Control *D. rerio* embryo at 72 h post fertilization (h.p.f.); the boxed region corresponds to images in the middle panels of **b**–**h**. **b**–**h**, Left: o-dianisidine staining was used to assess the number of mature erythrocytes at 48 h.p.f.: *ehd3* (**h**) morpholino-injected embryos showed normal haemoglobin staining, whereas embryos injected with *ak3* (**c**), *rnf145* (**d**), *arhgef3* (**e**) or *jmjd1c* (**g**) morpholinos showed a decrease in the number of haemoglobin-positive cells compared to control embryos (**b**). Embryos injected with *tpma* morpholinos (**f**) showed normal numbers of erythrocytes but unusual accumulation dorsally in the blood vessels (compatible with cardiomyopathy). Middle: haematopoietic stem-cell and thrombocyte development was assessed using the transgenic *Tg*(*cd41*:*EGFP*) line at 72 h.p.f. Embryos injected with the *ehd3* (**h**) morpholino had a normal number of GFP^+ cells in the caudal haematopoietic tissue and circulation, when compared to control embryos (**b**). However, GFP^+ cells were absent in *ak3* (**c**), *rnf145* (**d**), *arhgef3* (**e**), *tpma* (**f**) and *jmjd1c* (**g**) morpholino-injected embryos. Right: One-cell-stage embryos were injected with the standard control morpholino (**b**) or gene-specific morpholino (**c**–**h**) and monitored during development. No gross lethality or developmental abnormalities were observed at 72 h.p.f. in gene-specific morpholino-injected embryos (**c**–**h**) compared with the control (**b**). **a** and middle and right panels of **b**–**h**, lateral view, anterior left; left panels of **b**–**e**, **g**, **h**, ventral view, anterior up; left panel of **f**, dorsal view, anterior up. The genes appear to be nonspecifically expressed during embryogenesis as shown by patterns deposited in the ZFIN resource (http://zfin.org).

human homologue *SATB1* should be prioritized in functional studies. However, the results of the knockdown study in *D. rerio* do not clarify at which hierarchical positions in thrombopoiesis and erythropoiesis the genes exert their effect, requiring further assessment in conditional knockout models in *M. musculus* with lineage-specific regulation of gene transcription. Nevertheless, our results have already allowed novel insights into the genetic control of these processes. Signalling cascades initiated by thrombopoietin (THPO) and its receptor cMPL via the JAK2/STAT3/5A signalling pathway are key regulatory steps initiating changes in gene expression responsible for driving forward megakaryocyte differentiation[20]. Our study highlights several additional signalling proteins implicating potentially important novel regulatory routes. For instance, two genes encoding guanine nucleotide exchange factors (*DOCK8* and *ARHGEF3*) were identified. Mendelian mutations of the former are causative of the hyper-IgE syndrome, but its effect on platelets had not yet been identified. The silencing of the latter gene in *D. rerio* resulted in a profound haematopoietic phenotype characterized by a complete ablation of both primitive erythropoiesis and thrombocyte formation, demonstrating its novel regulatory role in myeloid differentiation. In a parallel and in-depth study we demonstrated its novel role in the regulation of iron uptake and erythroid cell maturation[21]. A second class of genes also known to critically control early and late events of megakaryopoiesis are transcription factors. For instance, *MYB* silencing by microRNA 150 determines the definitive commitment of the megakaryocyte–erythroblast precursor to the megakaryocytic lineage[15]. A further 10 core genes identified in this study are implicated in the regulation of transcription. Among these, we have demonstrated here that silencing of *rnf145* and *jmjd1c* in *D. rerio* severely affects both lineages.

In conclusion, this highly powered study describes a catalogue of known and novel genes associated with key haematopoietic processes in humans, providing an additional example of GWAS leading to biological discoveries. We further showed that for a large proportion of these known and new genes, functional support is achieved from model organisms and by overlap with genes implicated in inherited Mendelian disorders and in human cancers because of acquired mutations. In-depth functional studies and comparative analyses will be necessary to characterize the precise mechanisms by which these new genes and variants affect haematopoiesis, megakaryopoiesis and platelet formation. Furthermore, we provide extensive new resources, most notably a freely accessible knowledge base embedded in the novel protein–protein interaction network, with information about the identified platelet genes being implicated in Mendelian disorders and results from gene-silencing studies in model organisms. We anticipate that these resources will help to advance megakaryopoiesis research, to address key questions in blood stem-cell biology and to propose new targets for the treatment of haematological disorders. Finally, MPV has been associated with the risk of myocardial infarction[22,23]. The contribution of the new loci to the aetiology of acute myocardial infarction events will require assessment in a prospective setting.

METHODS SUMMARY

A summary of the methods can be found in Supplementary Information and includes detailed information on: study populations; blood biochemistry measurements; genotyping methods and quality control filters; genome-wide association and meta-analysis methods; gene prioritization strategies for functional assessment and network construction; protein–protein interaction network; *in vitro* differentiation of blood cells; experimental data sets and analytical methods for gene expression analysis; zebrafish morpholino knockdown generation; assessment of other model organism resources.

Received 17 December 2010; accepted 21 October 2011.
Published online 30 November 2011.

1. Frazer, K. A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449,** 851–861 (2007).

©2011 Macmillan Publishers Limited. All rights reserved

2. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55,** 997–1004 (1999).
3. Lo, K. S. *et al.* Genetic association analysis highlights new loci that modulate hematological trait variation in Caucasians and African Americans. *Hum. Genet.* **129,** 307–317 (2011).
4. Meisinger, C. *et al.* A genome-wide association study identifies three loci associated with mean platelet volume. *Am. J. Hum. Genet.* **84,** 66–71 (2009).
5. Soranzo, N. *et al.* A novel variant on chromosome 7q22.3 associated with mean platelet volume, counts, and function. *Blood* **113,** 3831–3837 (2009).
6. Soranzo, N. *et al.* A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nature Genet.* **41,** 1182–1190 (2009).
7. Bath, P. M. & Butterworth, R. J. Platelet size: measurement, physiology and vascular disease. *Blood Coagul. Fibrinolysis* **7,** 157–161 (1996).
8. Kamatani, Y. *et al.* Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nature Genet.* **42,** 210–215 (2010).
9. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461,** 747–753 (2009).
10. Park, J.-H. *et al.* Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nature Genet.* **42,** 570–575 (2010).
11. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467,** 832–838 (2010).
12. Teslovich, T. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466,** 707–713 (2010).
13. Paul, D. S. *et al.* Maps of open chromatin guide the functional follow-up of genome-wide association signals: application to hematological traits. *PLoS Genet.* **7,** e1002139 (2011).
14. Watkins, N. A. *et al.* A HaemAtlas: characterizing gene expression in differentiated human blood cells. *Blood* **113,** 1–9 (2009).
15. Lu, J. *et al.* MicroRNA-mediated control of cell fate in megakaryocyte-erythrocyte progenitors. *Dev. Cell* **14,** 843–853 (2008).
16. Crozatier, M. & Meister, M. *Drosophila* haematopoiesis. *Cell. Microbiol.* **9,** 1117–1126 (2007).
17. Fossett, N. *et al.* The Friend of GATA proteins U-shaped, FOG-1, and FOG-2 function as negative regulators of blood, heart, and eye development in *Drosophila*. *Proc. Natl Acad. Sci. USA* **98,** 7342–7347 (2001).
18. Goldfarb, A. N. Transcriptional control of megakaryocyte development. *Oncogene* **26,** 6795–6802 (2007).
19. Gregory, G. D. *et al.* FOG1 requires NuRD to promote hematopoiesis and maintain lineage fidelity within the megakaryocytic-erythroid compartment. *Blood* **115,** 2156–2166 (2010).
20. Wendling, F. *et al.* cMpl ligand is a humoral regulator of megakaryocytopoiesis. *Nature* **369,** 571–574 (1994).
21. Serbanovic-Canic, J. *et al.* Silencing of RhoA nucleotide exchange factor, ARHGEF3 reveals its unexpected role in iron uptake. *Blood* **118,** 4967–4976 (2011).
22. Chu, S. G. *et al.* Mean platelet volume as a predictor of cardiovascular risk: a systematic review and meta-analysis. *J. Thromb. Haemost.* **8,** 148–156 (2010).
23. Klovaite, J., Benn, M., Yazdanyar, S. & Nordestgaard, B. G. High platelet volume and increased risk of myocardial infarction: 39,531 participants from the general population. *J. Thromb. Haemost.* **9,** 49–56 (2010).
24. Garner, C. *et al.* Two candidate genes for low platelet count identified in an Asian Indian kindred by genome-wide linkage analysis: glycoprotein IX and thrombopoietin. *Eur. J. Hum. Genet.* **14,** 101–108 (2006).
25. Cline, M. S. *et al.* Integration of biological networks and gene expression data using Cytoscape. *Nature Protocols* **2,** 2366–2382 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Reactome is supported by grants from the US National Institutes of Health (P41 HG003751), EU grant LSHG-CT-2005-518254 'ENFIN', and the EBI Industry Programme. A.C. and D.L.S. are supported by the Wellcome Trust grants WT 082597/Z/07/Z and WT 077037/Z/05/Z and WT 077047/Z/05/Z, respectively. J.S.C. and K.V. are supported by the European Union NetSim training fellowship scheme (number 215820). A. Radhakrishnan, A.A., H.L.J., J.J., J.S. and W.H.O. are funded by the National Institute for Health Research, UK. Augusto Rendon is funded by programme grant RG/09/012/28096 from the British Heart Foundation. J.M.P. is supported by an Advanced ERC grant. IntAct is funded by the EC under SLING, grant agreement no. 226073 (Integrating Activity) within Research Infrastructures of the FP7, under PSIMEX, contract no. FP7-HEALTH-2007-223411. C.G. received support by the European Community's Seventh Framework Programme (FP7/2007-2013), ENGAGE Consortium, grant agreement HEALTH-F4-2007- 201413. N.S. is supported by the Wellcome Trust (Grant Number 098051). Acknowledgements by individual participating studies are provided in Supplementary Information.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.G. (christian.gieger@helmholtz-muenchen.de), W.H.O. (who1000@cam.ac.uk) or N.S. (ns6@sanger.ac.uk).

Author Contributions Study design group: C.G., S. Sanna, A.A.H., A. Rendon, M.A.F., W.H.O., N.S.; manuscript writing group: C.G., A. Radhakrishnan, S. Sanna, A.A.H., A. Rendon, M.A.F., W.H.O., N.S.; data preparation, meta-analysis and secondary analysis group: A. Radhakrishnan, B.K., W.T., E.P., G.P., R.M., M.A.F., C.G., N.S.; bioinformatics analyses, pathway analyses and protein–protein interaction network group: S.M., J.-W.N.A., S.J., J.K., Y.M., L.B., A. Rendon, W.H.O.; transcript profiling methods and data group: K.V., A. Rendon, L.W., A.H.G., T.-P.Y., F. Cambien, J.E., C. Hengstenberg, N.J.S., H. Schunkert, P.D., W.H.O.; *M. musculus* models: R.R.-S.; *D. rerio* knockdown models: A.C., J.S.-C., D. Stemple, W.H.O.; *D. melanogaster* knockdown models: U.E., J. Penninger and A.A.H. All other author contributions and roles are listed in Supplementary Information.

Christian Gieger[1]*, Aparna Radhakrishnan[2,3]*, Ana Cvejic[2,3]*, Weihong Tang[4]*, Eleonora Porcu[5]*, Giorgio Pistis[6]*, Jovana Serbanovic-Canic[2,3]*, Ulrich Elling[7], Alison H. Goodall[8,9], Yann Labrune[10], Lorna M. Lopez[11,12], Reedik Mägi[13,28], Stuart Meacham[2,3], Yukinori Okada[14,15], Nicola Pirastu[16,102], Rossella Sorice[17], Alexander Teumer[18], Katrin Voss[2], Weihua Zhang[19], Ramiro Ramirez-Solis[3], Joshua C. Bis[20], David Ellinghaus[21], Martin Gögele[22], Jouke-Jan Hottenga[23], Claudia Langenberg[24], Peter Kovacs[25], Paul F. O'Reilly[26], So-Youn Shin[3,27], Tõnu Esko[28,29,30], Jaana Hartiala[31,32], Stavroula Kanoni[3,33], Federico Murgia[34], Afshin Parsa[35], Jonathan Stephens[2], Pim van der Harst[36], C. Ellen van der Schoot[37], Hooman Allayee[31,32], Antony Attwood[2,3], Beverley Balkau[38,39], François Bastardot[40], Saonli Basu[41], Sebastian E. Baumeister[42], Ginevra Biino[34,43,44], Lorenzo Bomba[45], Amélie Bonnefond[10], François Cambien[46], John C. Chambers[19], Francesco Cucca[5], Pio D'Adamo[16,102], Gail Davies[12], Rudolf A. de Boer[36], Eco J. C. de Geus[23], Angela Döring[47,48], Paul Elliott[19,49], Jeanette Erdmann[50,101], David M. Evans[51], Mario Falchi[52], Wei Feng[41], Aaron R. Folsom[4], Ian H. Frazer[53], Quince D. Gibson[35], Nicole L. Glazer[54], Chris Hammond[27], Anna-Liisa Hartikainen[55], Susan R. Heckbert[56,57], Christian Hengstenberg[58], Micha Hersch[59], Thomas Illig[60], Ruth J. F. Loos[24], Jennifer Jolley[2], Kay Tee Khaw[61,62], Brigitte Kühnel[1], Marie-Christine Kyrtsonis[63], Vasiliki Lagou[64], Heather Lloyd-Jones[2], Thomas Lumley[65], Massimo Mangino[27], Andrea Maschio[5], Irene Mateo Leach[36], Barbara McKnight[65], Yasin Memari[3,27], Braxton D. Mitchell[35], Grant W. Montgomery[66], Yusuke Nakamura[67], Matthias Nauck[68], Gerjan Navis[69], Ute Nöthlings[70,71], Ilja M. Nolte[72], David J. Porteous[11,73], Anneli Pouta[55,74], Peter P. Pramstaller[22], Janne Pullat[28], Susan M. Ring[51], Jerome I. Rotter[75], Daniela Ruggiero[17], Aimo Ruokonen[76], Cinzia Sala[6], Nilesh J. Samani[8,9], Jennifer Sambrook[2,77], David Schlessinger[78], Stefan Schreiber[21], Heribert Schunkert[50,101], James Scott[79], Nicholas L. Smith[56,57,80], Harold Snieder[72], John M. Starr[11,81], Michael Stumvoll[82], Atsushi Takahashi[14], W. H. Wilson Tang[83,84], Kent Taylor[75], Albert Tenesa[85,86], Swee Lay Thein[87], Anke Tönjes[82], Manuela Uda[5], Sheila Ulivi[16], Dirk J. van Veldhuisen[36], Peter M. Visscher[11,66], Uwe Völker[18], H.-Erich Wichmann[47,88], Kerri L. Wiggins[20], Gonneke Willemsen[23], Tsun-Po Yang[3], Jing Hua Zhao[24], Paavo Zitting[89], John R. Bradley[77], George V. Dedoussis[33], Paolo Gasparini[16,102], Stanley L. Hazen[83], Andres Metspalu[28,29,30], Mario Pirastu[34,43], Alan R. Shuldiner[35,90], L. Joost van Pelt[91], Jaap-Jan Zwaginga[92], Dorret I. Boomsma[23], Ian J. Deary[11,12], Andre Franke[21], Philippe Froguel[10,52], Santhi K. Ganesh[93], Marjo-Riitta Jarvelin[26,94,95], Nicholas G. Martin[66], Christa Meisinger[48], Bruce M. Psaty[57,96], Timothy D. Spector[27], Nicholas J. Wareham[24], Jan-Willem N. Akkerman[97], Marina Ciullo[17], Panos Deloukas[3], Andreas Greinacher[98], Steve Jupe[99], Naoyuki Kamatani[14], Jyoti Khadake[99], Jaspal S. Kooner[79], Josef Penninger[7], Inga Prokopenko[64], Derek Stemple[3], Daniela Toniolo[6,44], Lorenz Wernisch[100], Serena Sanna[5]*, Andrew A. Hicks[22]*, Augusto Rendon[2,100]*, Manuel A. Ferreira [66]*, Willem H. Ouwehand [2,3,77]* & Nicole Soranzo[3]*

[1]Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstr.1, 85764 Neuherberg, Germany. [2]Department of Haematology, University of Cambridge & NHS Blood and Transplant, Cambridge CB2 0PT, UK. [3]Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1HH/1SA, UK. [4]Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, Minnesota 55455-0354, USA. [5]Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, c/o Cittadella Universitaria di Monserrato, Monserrato, Cagliari 09042, Italy. [6]San Raffaele Scientific Institute, Via Olgettina 58, 20132 Milano, Italy. [7]IMBA, Institute of Molecular Biotechnology of the Austrian Academy of Sciences, 1030 Vienna, Austria. [8]Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Groby Road, Leicester LE3 9QP, UK. [9]Leicester NIHR Biomedical Research Unit in Cardiovascular Disease, Glenfield Hospital, Leicester LE3 9QP, UK. [10]CNRS-UMR-8199, Institut de Biologie de Lille, 1 Rue du Professor Calmette, BP245, 59019 Lille Cedex, France. [11]Centre for Cognitive Ageing and Cognitive Epidemiology, 7 George Square, The University of Edinburgh, Edinburgh EH8 9JZ, UK. [12]Department of Psychology, The University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK. [13]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. [14]Laboratory for Statistical Analysis, RIKEN Center for Genomic Medicine, Institute of Physical and Chemical Research, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. [15]Department of Allergy and Rheumatology, Graduate School of Medicine, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8654, Japan. [16]Institute for Maternal and Child Health — IRCCS 'Burlo Garofolo' — via dell'Istria 65/1, 34137 Trieste, Italy. [17]Institute of Genetics and Biophysics A. Buzzati-Traverso, CNR, Via P. Castellino 111, 8031 Naples, Italy. [18]Interfaculty Institute for Genetics and Functional Genomics, Ernst-Moritz-Arndt-University Greifswald, 17487 Greifswald, Germany. [19]Department of Epidemiology and Biostatistics, School of Public Health, Faculty of Medicine, St Mary's Campus, London W2 1PG, UK. [20]Department of Medicine, University of Washington, 1959 Northeast Pacific Street, Seattle, Washington 98195-0001, USA. [21]Institute of Clinical Molecular Biology, Christian-Albrechts-University, 24105 Kiel, Germany. [22]Centre for Biomedicine, European Academy Bozen/Bolzano, Viale Druso 1, 39100 Bolzano, Italy. [23]Department of Biological Psychology, VU University Amsterdam, De Boelelaan 1105, Van der Boechorststraat 1, 1081 BT, Amsterdam, The Netherlands. [24]MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Box 285, Cambridge CB2 0QQ, UK. [25]Interdisciplinary Centre for Clinical Research, University of Leipzig, Inselstr. 22, 04103 Leipzig, Germany. [26]Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK. [27]Department of Twin Research and Genetic

©2011 Macmillan Publishers Limited. All rights reserved

Epidemiology, King's College London, London SE1 7EH, UK. [28]Estonian Genome Center, University of Tartu, Tiigi 61b, 50410 Tartu, Estonia. [29]Estonian Biocenter, Riia 23b, 51010, Tartu, Estonia. [30]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. [31]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California 90089-9075, USA. [32]Institute for Genetic Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California 90089-9075, USA. [33]Department of Nutrition-Dietetics, Harokopio University, Athens 17671, Greece. [34]Institute of Population Genetics, National Council of Research, 07100 Sassari, Italy. [35]University of Maryland School of Medicine, 655 West Baltimore Street, Baltimore, Maryland 21201-1559, USA. [36]Department of Cardiology, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713 GZ Groningen, The Netherlands. [37]Department of Experimental Immunohaematology, Sanquin Research, Plesmanlaan 125, 1066 CX Amsterdam, The Netherlands and Landsteiner Laboratory, Academic Medical Centre, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands. [38]Inserm, CESP Centre for research in Epidemiology and Population Health, U1018 Villejuif, 94807 France. [39]University Paris Sud 11, UMRS 1018, Villejuif, 94807 France. [40]Service de Médecine Interne, CHUV, 1011 Lausanne, Switzerland. [41]Division of Biostatistics, School of Public Health, University of Minnesota, A460 Mayo Building, mmc 303, 420 Delaware St SE, Minneapolis, Minnesota 55455, USA. [42]Institute for Community Medicine, Ernst-Moritz-Arndt-University Greifswald, 17475 Greifswald, Germany. [43]Shardna Life Sciences, Loc. Piscinamanna 09010 Pula (CA), Italy. [44]Institute of Molecular Genetics-CNR Via Abbiategrasso 207, 27100 Pavia, Italy. [45]Istituto di Zootecnica, Universita' Cattolica del Sacro Cuore sede Piacenza, via Emilia Parmense 84, 29122 Piacenza, Italy. [46]INSERM UMRS 937, Pierre and Marie Curie University (UPMC, Paris 6) and Medical School, 91 Bd de l'Hôpital, 75013 Paris, France. [47]Institute of Epidemiology I, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstr.1, 85764 Neuherberg, Germany. [48]Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstr.1, 85764 Neuherberg, Germany. [49]MRC-HPA Centre for Environment and Health, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PG, UK. [50]Universität zu Lübeck, Medizinische Klinik 2, Lübeck Germany. [51]MRC Centre for Causal Analyses in Translational Epidemiology, School of Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK. [52]Department of Genomics of Common Disease, School of Public Health, Imperial College London, The Commonwealth Building, The Hammersmith Hospital, Du Care Road, London W12 0NN, UK. [53]The University of Queensland Diamantina Institute, Princess Alexandra Hospital, Brisbane, Queensland 4102, Australia. [54]Department of Medicine, Boston University, Boston, Massachusetts 02118, USA. [55]Department of Clinical Sciences/Obstetrics and Gynecology, University of Oulu, Aapistie 7, Oulu, 90220 Finland. [56]Department of Epidemiology, University of Washington, Seattle, Washington 98195-7236, USA. [57]Group Health Research Institute, Group Health Cooperative, Seattle, Washington 98101-1448, USA. [58]Klinik und Poliklinik für Innere Medizin II, Universitätsklinikum Regensburg, 93053 Regensburg, Germany. [59]Department of Medical Genetics, University of Lausanne, and Swiss Institute of Bioinformatics, Lausanne, Switzerland. [60]Unit for Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstr.1, 85764 Neuherberg, Germany. [61]Department of Public Health and Primary Care, Strangeways Research Laboratory, University of Cambridge, Cambridge CB1 8RN, UK. [62]Clinical Gerontology Unit, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, UK. [63]First Department of Propedeutic Internal Medicine and Hematology Department, National and Kapodistrian University of Athens Medical School, Laikon General Hospital, Athens 11727, Greece. [64]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford OX3 7LJ, UK. [65]Department of Biostatistics, University of Washington, Seattle, Washington 98195-7232, USA. [66]Genetic Epidemiology, Queensland Institute of Medical Research, The Bancroft Centre, 300 Herston Road, Herston 4006, Queensland, Australia. [67]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan. [68]Institute of Clinical Chemistry and Laboratory Medicine, Ernst-Moritz-Arndt-University Greifswald, 17475 Greifswald, Germany. [69]Department of Internal Medicine, University Medical Center Groningen, University of Groningen, 9713 GZ Groningen, The Netherlands. [70]Section for Epidemiology, Institute for Experimental Medicine, Christian-Albrechts-University Kiel, 24105 Kiel, Germany. [71]Popgen Biobank, University Clinic Schleswig-Holstein, 24105 Kiel, Germany. [72]Unit of Genetic Epidemiology and Bioinformatics, Department of Epidemiology, University Medical Center Groningen, University of Groningen, 9713 GZ Groningen, The Netherlands. [73]Medical Genetics Section, University of Edinburgh, Molecular Medicine Centre, Western General Hospital, Crewe Road South, Edinburgh EH4 2XU, UK. [74]National Institute for Health and Welfare, Aapistie 1, Oulu 90220, Finland. [75]Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California 90048, USA. [76]Institute of Diagnostics, Clinical Chemistry, University of Oulu, Box 5000Y, University of Oulu, Oulu 90014, Finland. [77]NIHR Cambridge Biomedical Research Centre, Cambridge CB2 0QQ, UK. [78]Laboratory of Genetics, National Institute on Aging, Baltimore, Maryland 21224-6825, USA. [79]Faculty of Medicine, National Heart and Lung Institute, Cardiovascular Science, Hammersmith Campus, Hammersmith Hospital, Imperial College London, London W12 0HS, UK. [80]Seattle Epidemiologic Research and Information Center of the Department of Veterans Affairs Office of Research and Development, Seattle, Washington 98108, USA. [81]Geriatric Medicine Unit, The University of Edinburgh, Royal Victoria, Craigleith Road, Edinburgh EH4 2DN, UK. [82]Department of Medicine, University of Leipzig, Liebigstr. 18, 04103 Leipzig, Germany. [83]Center for Cardiovascular Diagnostics and Prevention, Department of Cell Biology and Cardiovascular Medicine, Cleveland Clinic, Cleveland, Ohio 44195, USA. [84]Heart and Vascular Institute, Cleveland Clinic, Cleveland, Ohio 44195, USA. [85]Institute of Genetics and Molecular Medicine, MRC Human Genetics Unit, Edinburgh EH4 2XU, UK. [86]The Roslin Institute, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Roslin Midlothian EH25 9RG, UK. [87]Department of Molecular Haematology, King's College London, 125 Coldharbour Lane, London SE5 9NU, UK. [88]Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität and Klinikum Grosshadern, Marchioninistr. 15, 81377 Munich, Germany. [89]Department of Physiatrics, Lapland Central Hospital, Rovaniemi, 96101, Finland. [90]Geriatric Research and Education Clinical Center, Veterans Administration Medical Center, Baltimore, Maryland 21208, USA. [91]Department of Laboratory Medicine, University Medical Center Groningen, University of Groningen, 9713 GZ, Groningen, The Netherlands. [92]Sanquin Research Leiden, Jon J. van Rood Center for Clinical Transfusion research, and the Department of Immuno-hematology and Blood Transfusion, Leiden University Medical Center, 2333 Leiden, The Netherlands. [93]Division of Cardiovascular Medicine, University of Michigan Health System, Ann Arbor, Michigan, USA. [94]Institute of Health Sciences, University of Oulu, Aapistie 7, Oulu, 90220 Finland. [95]National Institute for Health and Welfare, Biocenter Oulu, University of Oulu, Aapistie 7, Oulu, 90220 Finland. [96]Department of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, Washington 98101-1448, USA. [97]Department of Clinical Chemistry and Haematology, University Medical Center Utrecht, 3584 CX Utrecht, Netherlands. [98]Institute for Immunology and Transfusion Medicine, Ernst-Moritz-Arndt-University Greifswald, 17475 Greifswald, Germany. [99]Proteomics Services, EMBL–European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. [100]MRC Biostatistics Unit, Institute of Public Health, University Forvie Site, Robinson Way, Cambridge CB2 0SR, UK. [101]Nordic Center of Cardiovascular Research, Lubeck, Germany. [102]University of Trieste, Piazzale Europa 1, 34127 Trieste, Italy.
*These authors contributed equally to this work.

©2011 Macmillan Publishers Limited. All rights reserved

ARTICLE

doi:10.1038/nature10649

The circadian molecular clock creates epidermal stem cell heterogeneity

Peggy Janich[1], Gloria Pascual[1], Anna Merlos-Suárez[2], Eduard Batlle[2,3], Jürgen Ripperger[4], Urs Albrecht[4], Hai-Ying M. Cheng[5], Karl Obrietan[6], Luciano Di Croce[1,3] & Salvador Aznar Benitah[1,3]

Murine epidermal stem cells undergo alternate cycles of dormancy and activation, fuelling tissue renewal. However, only a subset of stem cells becomes active during each round of morphogenesis, indicating that stem cells coexist in heterogeneous responsive states. Using a circadian-clock reporter-mouse model, here we show that the dormant hair-follicle stem cell niche contains coexisting populations of cells at opposite phases of the clock, which are differentially predisposed to respond to homeostatic cues. The core clock protein Bmal1 modulates the expression of stem cell regulatory genes in an oscillatory manner, to create populations that are either predisposed, or less prone, to activation. Disrupting this clock equilibrium, through deletion of *Bmal1* (also known as *Arntl*) or *Per1/2*, resulted in a progressive accumulation or depletion of dormant stem cells, respectively. Stem cell arrhythmia also led to premature epidermal ageing, and a reduction in the development of squamous tumours. Our results indicate that the circadian clock fine-tunes the temporal behaviour of epidermal stem cells, and that its perturbation affects homeostasis and the predisposition to tumorigenesis.

Epidermal stem cells ensure that skin homeostasis is maintained. Murine epidermal stem cells are located either at the permanent portion of the hair follicle—termed the bulge—and are exclusively responsible for hair cycling[1–4]; or at the junction between the epidermis and the hair follicle (isthmus), and feed into the epidermis and sebaceous glands[5–7]. In addition, a continuous proliferation of basal interfollicular epidermal cells ensures daily epidermal maintenance[8].

Bulge stem cells undergo bouts of activation followed by periods of dormancy, to establish hair follicle cycling. Robust TGF-β and Bmp signals act as 'activation breaks', rendering bulge cells dormant during the resting phase of the hair cycle (telogen)[9–11]. At the onset of the growth phase (anagen), bulge cells respond to Wnt signals by migrating into the lower proliferative hair germ region, where they contribute to follicle growth[12–15]. Subsequently, at mid-anagen, the bulge undergoes a second round of activation, which replenishes cells lost at the onset of anagen[2,3]. However, the response of bulge stem cells to activating stimuli is a heterogeneous process, as only a subset of them become active during either stage of activation[12,13]. The nature of such niche heterogeneity is currently unknown. Importantly, perturbing the equilibrium between the responsive and non-responsive stem cell states causes tissue malfunction and increases the risk of carcinogenesis[16–20].

Here, we analysed the role of the molecular clock in fine-tuning the function of epidermal stem cells. The mammalian clock machinery anticipates and synchronizes vital functions related to the physiological circadian needs of the organism[21]. The core molecular clock is established by a positive limb, composed of heterodimers of the transcription factors Clock and Bmal, which drives the rhythmic expression of the negative limb proteins, period (Per1–3), cryptochrome (Cry1/2), and Dec1/2 (ref. 21). Accumulated Per and Cry proteins in turn translocate to the nucleus and inhibit Bmal/Clock transcriptional activity, thereby repressing their own expression and marking the beginning of a new cycle.

Circadian heterogeneity in hair-follicle stem cells

We monitored the activity of the clock in epidermal stem cells by means of a reporter bacterial artificial chromosome (BAC) transgenic mouse, in which the expression of fluorescent venus is under the regulation of the full-length promoter of the *Per1* gene[22]. Venus expression mirrors the endogenous oscillation of the clock in the suprachiasmatic nuclei, where the central pacemaker resides, thereby establishing its bona fide circadian reporter activity[22]. We first studied the behaviour of the clock in the dorsal skin of Per1–venus mice collected between postnatal days (P)19 and P31, when hair follicles synchronously transit from the dormant to the growth phase.

At P19, a stage in which the bulge is predominantly dormant, the bulge cells expressing CD34 and the highest levels of α6 integrin (α6 integrin^{bright}/CD34^{+}) contained a continuum of venus-expressing populations, which could be subdivided into cells with highest mean fluorescence intensity of venus (venus^{bright}), and those with no venus fluorescence (venus^{dim}), as determined by immunohistochemistry (Supplementary Fig. 1), and fluorescence-activated cell sorting (FACS) (Fig. 1a, b). At this stage, the bulge contained approximately equal numbers of venus^{bright} and venus^{dim} stem cells (Fig. 1a, b). As hair follicles synchronously progressed into anagen (from P20 to P31), the proportion of venus^{bright} bulge cells steadily increased, reaching a proportion of 90% venus^{bright} to 10% venus^{dim} at the peak of follicle growth (Fig. 1a, b). Conversely, the basal layer of the interfollicular epidermis (α6 integrin^{bright}/CD34^{-}) showed a homogenous pattern of clock activity with a ratio of 95:5 venus^{bright}:venus^{dim} cells, irrespective of the day analysed (Supplementary Fig. 2). Heterogeneity of the clock phase in bulge cells during the telogen to anagen transition was also observed with another independently generated reporter line, Per1–GFP, as shown by three-dimensional whole-mount GFP fluorescence imaging of tail epidermis (Supplementary Fig. 3a, b)[23].

We next verified whether these clock stem cells states showed circadian rhythmicity. FACS analysis indicated that the expression of venus in bulge stem cells in anagen (P27), or interfolllicular epidermis basal cells, followed a circadian pattern, irrespective of whether the mice were maintained in 12 h light/dark cycles, or in constant darkness under free running conditions (Fig. 1c and Supplementary Fig. 4a, b)[21]. This circadian variation was further confirmed

[1]Center for Genomic Regulation and UPF, 08003 Barcelona, Spain. [2]Institute for Research in Biomedicine, 08028 Barcelona, Spain. [3]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain. [4]University of Fribourg, 1700 Fribourg, Switzerland. [5]Department of Biology, University of Toronto Mississauga, 3359 Mississauga Road N., Mississauga, Ontario L5L 1C6, Canada. [6]Ohio State University, Columbus, Ohio 43210, USA.

©2011 Macmillan Publishers Limited. All rights reserved

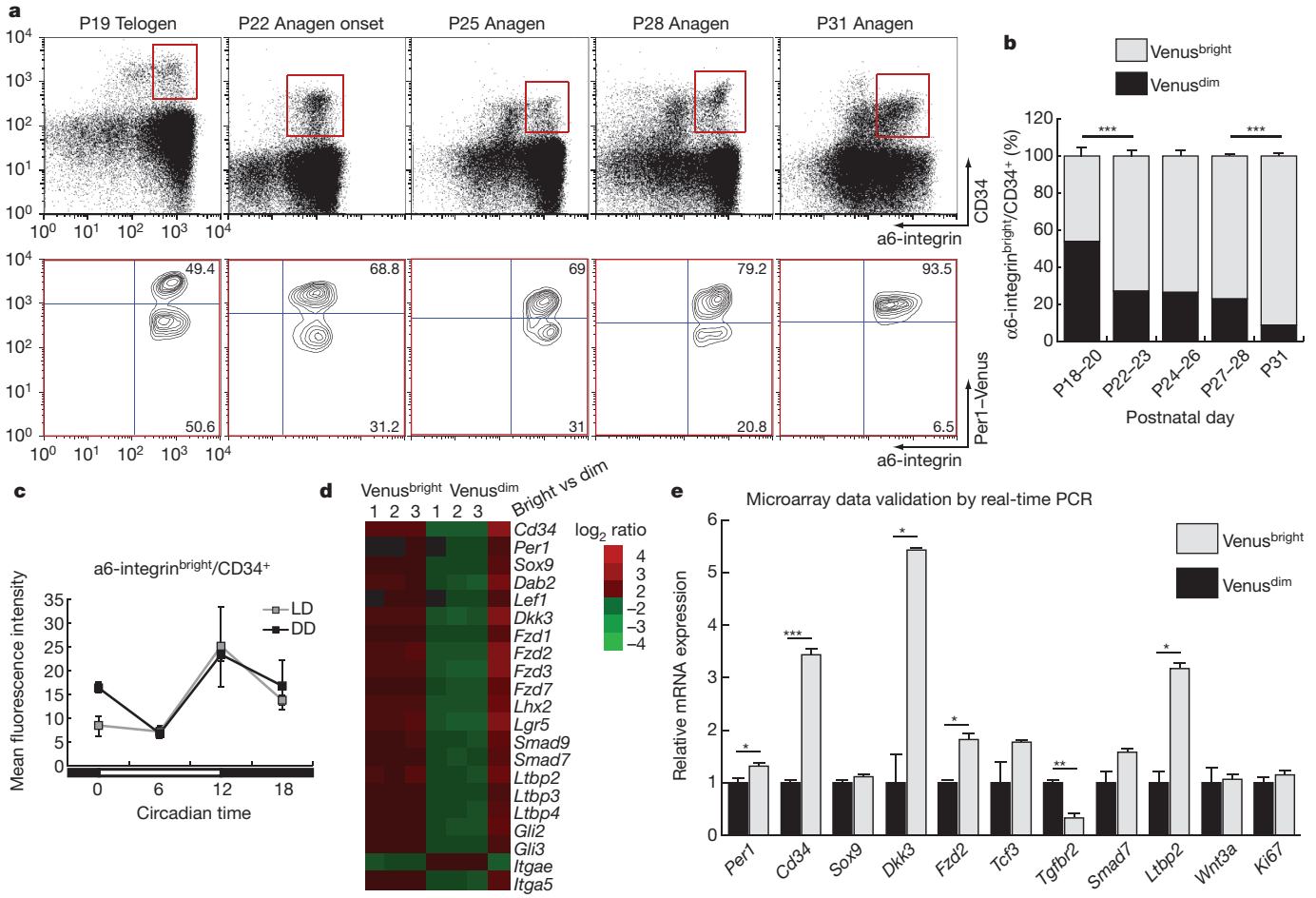


Figure 1 | The molecular clock regulates the expression of the bulge stem cell signature. **a**, **b**, Percentages of venus^{bright} and venus^{dim} cells in the bulge by FACS ($n \geq 6$). **c**, Venus mean fluorescence intensity of bulge cells in mice kept under 12 h light/12 h dark (LD) and 12 h dark/12 h dark (DD) conditions ($n = 2$). **d**, Heatmap of selected genes from arrays of venus^{bright} and venus^{dim} bulge cells of P19 mice ($n = 3$). **e**, Validation of microarray data by real-time PCR. Fold change is shown as relative to venus^{dim} cells after normalization to pumilio 1 (*Pum1*) ($n = 2$, pool of 6 mice per replicate). Results in **b**, **c** and **e** are shown as mean ± s.e.m., $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ (two-tailed Student's *t*-test).

by time-lapse *in vivo* confocal microscopy of venus fluorescence in dorsal skin explants biopsied from adult Per1–venus mice, and GFP fluorescence in whole mounts of tail epidermis of Per1–GFP mice (Supplementary Fig. 4c, d).

The clock regulates stem cell genes

We then compared the global transcriptomes of purified venus^{bright} and venus^{dim} bulge cells from the dorsal skin of P18–19 mice (α6^{bright}/CD34^{+}/venus^{bright} and α6^{bright}/CD34^{+}/venus^{dim}). As expected, both populations showed differential expression of core circadian transcripts, such as *Cry2*, *Per1*, *Nr1d1*, *Rorβ*, *Dec2* and *E4BP4* (also known as *Nfil3*) (Fig. 1d and Supplementary Table 1). Intriguingly, although the bulge is inactive at P18–19, both populations differed in the expression of a significant number of genes previously shown to constitute the bulge signature[2,3,12,24]. Venus^{bright} bulge stem cells expressed higher levels (between 1.4- to 3-fold) of Wnt-signalling factors, including *Tcf3*, *Fzd2/3*, *Sox9*, *Lhx2*, *Lgr5*, *Lef1*, *Dkk3* and *Dab2*, as well as TGF-β-inhibitory factors such as *Smad7*, *Ltbp2–4*, *Smurf*, *Lefty* and *Cul1* (Fig. 1d, e, and Supplementary Table 1). Other pathways differentially expressed relevant for bulge behaviour related to integrins, Notch, Bmp and Shh, among others (Fig. 1d, e).

Thus, the coexisting clock states of dormant bulge stem cells correlated with differential expression of key epidermal homeostasis genes. Promoter analysis revealed that several of these genes, encoding for proteins known to control bulge dormancy, activation, differentiation and niche interactions[1], contained several putative Bmal1/Clock-binding sites within their proximal and distal promoter regions (Supplementary Table 2). These included the Wnt signalling factors *Dab2*, *Lef1*, *Dkk3*, *Fzd2*, *Sox9*, *Lhx2* and *Tcf4*; TGF-β regulators such as *Smad7*, *Lefty*, *Smurf2* and *Smad9*; and *Itga6* as well as modulators of Bmp and Notch signalling (Supplementary Table 2). We confirmed by chromatin immunoprecipitation (ChIP) that Bmal1/Clock bound to these promoters in intact adult tail epidermis (Supplementary Fig. 5a), and that the binding of Bmal1 was circadian (Fig. 2). Chromatin occupancy of Bmal1 to these genes was also confirmed in FACS-sorted bulge stem cells (Supplementary Fig. 5b).

Clock arrhythmia affects homeostasis

We next sought to study the biological significance of this clock mechanism in epidermal stem cells *in vivo*. To this end, we generated mice with a conditional deletion of *Bmal1* in the keratin-14^{+} basal keratinocyte compartment (K14Cre/Bmal1$^{LoxP/LoxP}$, hereafter refered as Bmal1KO) (Supplementary Fig. 6a, b). Deletion of *Bmal1* causes circadian arrhythmicity without the need to perturb any other core circadian member[21], and its ubiquitous deletion causes premature ageing, including defects in adult skin morphogenesis[25–27]. We crossed Bmal1KO and Per1–venus mice, and verified that the circadian clock of bulge stem cells and basal interfollicular epidermal cells was arrhythmic, and permanently skewed towards a clock^{low} (venus^{dim}) state (Supplementary Fig. 6c).

Bulge stem cells and primary keratinocytes isolated from the dorsal skin of Bmal1KO mice expressed lower transcript levels of Wnt-related

©2011 Macmillan Publishers Limited. All rights reserved

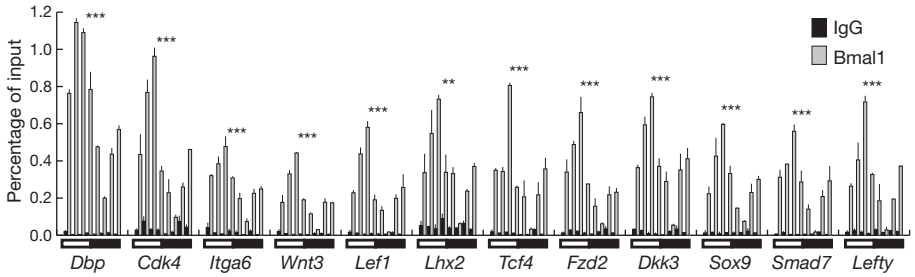


Figure 2 | Circadian binding of Bmal to the promoters of genes involved in adhesion, cell cycle, TGF-β and Wnt pathways. ChIP from tail epidermis of wild-type mice collected every 3 h during 24 h (white and black bars represent day and night, respectively). Graph shows percentage of immunoprecipitated DNA over an input control from one representative experiment ($n = 2$). Results are shown as mean ± s.e.m., $^{**}P < 0.01$, $^{***}P < 0.001$ (one-way ANOVA). Statistical significance determined by Cosinor analysis revealed a P value of $P < 0.05$ for *Dbp* and *Cdk4*, and $P < 0.001$ for *Itga6*, *Lef1*, *Smad7*, *Sox9* and *Tcf4*.

genes including *Dab2*, *Dkk3*, *Lef1* and *Wnt10a*, than control stem cells (Fig. 3a and Supplementary Fig. 5c). In addition, they contained lower messenger RNA levels of TGF-β inhibitors, and higher amounts of *Tgfbr2* and *Smad3* (Fig. 3a and Supplementary Fig. 5c). The expression of Wnt and TGF-β-related factors in purified bulge stem cells varied within a 12 h period in wild-type mice, but not Bmal1KO mice (Supplementary Fig. 5d). Accordingly, the hair follicles of Bmal1KO mice showed the same differences at the protein level, as exemplified by immunohistochemical analysis of Sox9, Lef1, phospho-Smad2 and TGFβR2 in dorsal skin sections, and western blot from primary keratinocytes (Fig. 3b, d).

Altogether, these results indicated that the clock machinery might endow subpopulations of epidermal stem cells with different predispositions to respond to dormancy or activation stimuli, such as TGF-β and Wnt. In line with this, the hair follicle bulge and interfollicular epidermis of Bmal1KO mice contained higher levels of active phospho-Smad2 (Fig. 3b, d). Moreover, Bmal1KO keratinocytes were more responsive to TGF-β treatment than control keratinocytes (Fig. 3c). We could not detect any Wnt activity using the Wnt-specific reporter containing binding sites for TCF/Lef proteins (TOP-Flash) in our primary cultures of mouse keratinocytes stimulated with the GSK3b inhibitor 6-bromoindirubin-3-oxime (BIO), or purified Wnt3a, in accordance with previous reports[28]; this prevented us from further studying the effect of Bmal1 deletion on Wnt responsiveness (Supplementary Fig. 7).

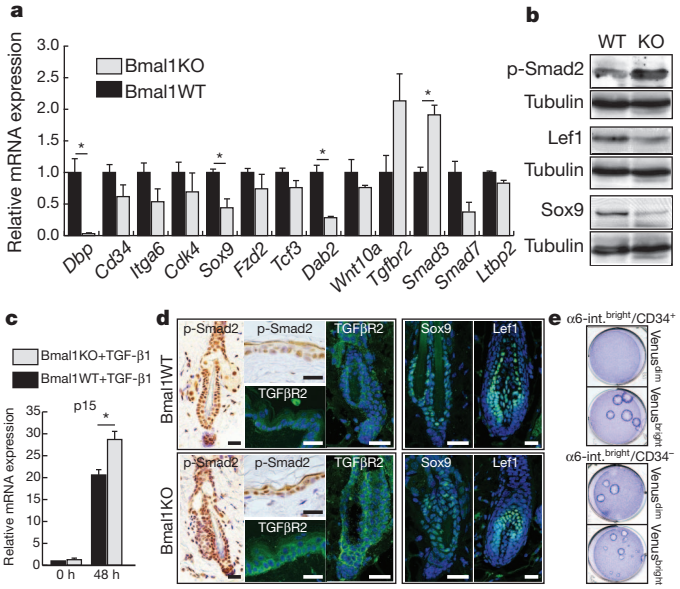


Figure 3 | Bmal1 modulates the response of bulge stem cells to activation and dormancy cues. a, Differential expression of genes in bulge cells of Bmal1WT and Bmal1KO mice. Fold-change values are shown as relative expression to Bmal1WT cells after normalization to *Pum1* ($n = 2$). **b**, Western blot analysis for phospho-Smad2 (p-Smad2), Lef1 and Sox9 in tail epidermis of 9-month-old Bmal1WT and Bmal1KO mice; $n = 3$ mice were analysed for each group; **c**, Primary mouse keratinocytes of Bmal1KO show enhanced responsiveness to TGF-β1 after 48 h of treatment ($n = 3$). **d**, Immunostaining for phospho-Smad2, TGFβR2, Lef1 and Sox9 in back skin of Bmal1WT and Bmal1KO mice ($n \geq 5$). Scale bars 25 μm. **e**, Clonogenic assay of FACS-purified venus^{bright} and venus^{dim} bulge and interfollicular epidermis (IFE) keratinocytes from the back skin of P19 Per1–venus mice (7×10^3 bulge and 1×10^5 epidermal cells). Results in **a** and **c** are shown as mean ± s.e.m., $^{*}P < 0.05$ (two-tailed Student's t-test).

When plated at clonal density, Per1–venus^{bright} bulge stem cells and basal epidermal cells showed a higher growth potential than the corresponding venus^{dim} population, further suggesting that the clock^{high} state is more prone to become activated than the clock^{low} counterpart (Fig. 3e). Additional data confirmed this hypothesis. First, the hair follicle bulges of Bmal1KO mice, which are permanently locked in the clock^{low} state, contained fewer proliferative cells and a higher number of long-term label (BrdU)-retaining dormant stem cells (LRCs), from 10 months of age and onwards (Fig. 4a). Second, the epidermis of another model of circadian arrhythmia, Per1 and Per2 double-mutant mice (Per1/2dKO), which lacks the negative limb of the molecular clock and is therefore locked in the clock^{high} state, showed the opposite effects; that is, enhanced bulge proliferation, reduced numbers of bulge LRCs and sustained expression of epidermal clock target genes[29] (Fig. 4b and Supplementary Fig. 8). The hair follicles of Bmal1KO were less efficient in becoming active upon depilation (Supplementary Fig. 9a). Bmal1KO bulge cells were also less hyperprolifertive than wild-type bulge cells in response to treatment with the phorbol ester 12-O-tetradecanoylphorbol-13-acetate (TPA), thus delaying the entry of the hair follicles into anagen (Supplementary Fig. 9b). Lastly, epidermal stem cells purified from Bmal1KO and Per1/2dKO mice were less and more clonogenic, respectively, *in vitro* than wild-type cells (Supplementary Fig. 10a, b). Of note, deletion of *Bmal1* or *Per1/2* did not affect the proportion of bulge stem cells in adulthood compared to their controls (Supplementary Fig. 11a, b).

Loss of Bmal1 induces epidermal ageing

Reduced and enhanced proliferation, respectively, was also evident in the basal layer of the interfollicular epidermis of Bmal1KO and Per1/2dKO mice (Supplementary Fig. 12). Bmal1KO mice showed signs of inefficient epidermal self-renewal, with premature signs of ageing as early as 5 months of age, such as the accumulation of terminally differentiated cornified cells (Fig. 4c and Supplementary Fig. 9c). This was accompanied by increased expression of p16, which has been previously associated with increased epidermal ageing[30], but not p19 or apoptosis (Supplementary Fig. 13). Because bulge stem cells do not contribute to epidermal maintenance in steady-state conditions[31], we sought to understand the molecular mechanisms underlying the defects of the interfollicular epidermis of Bmal1KO mice. We performed

©2011 Macmillan Publishers Limited. All rights reserved

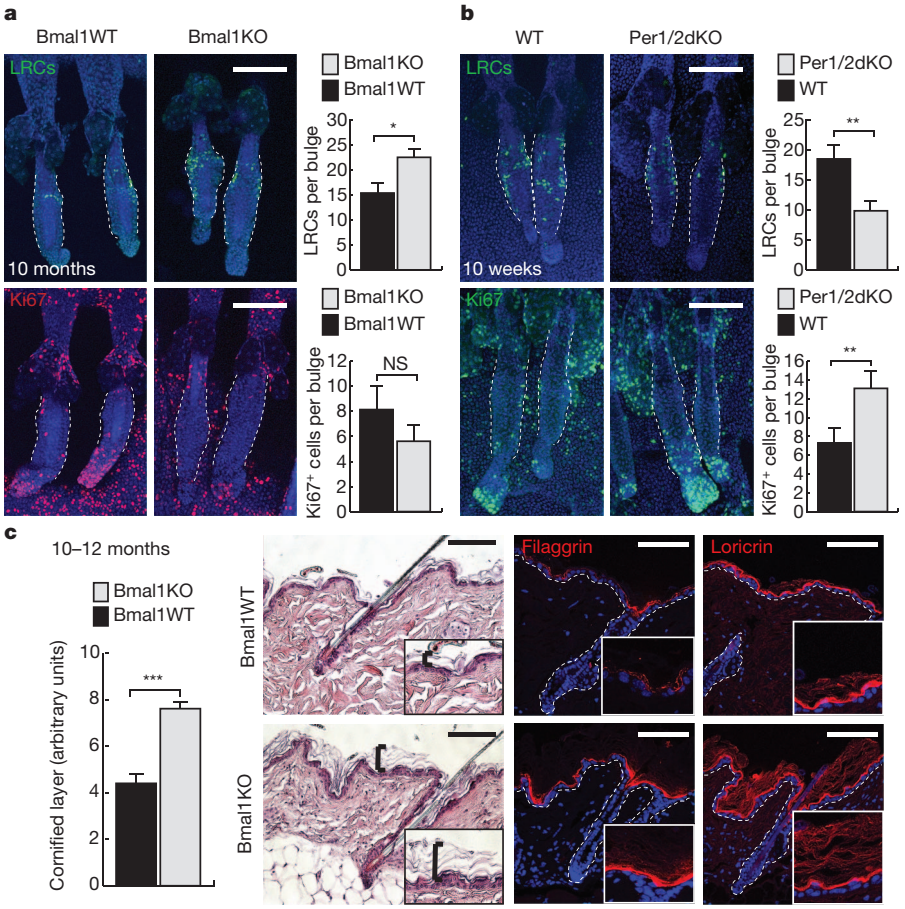


Figure 4 | Clock perturbation *in vivo* results in changes in the number of dormant bulge stem cells, and premature epidermal ageing.
a, **b**, Quantification of BrdU$^+$ LRCs and Ki67$^+$ cells in the bulge of 10-month-old Bmal1WT and Bmal1KO mice (**a**) and 10-week-old wild type (WT) and Per1/2dKO mice (**b**), showing opposite phenotypes ($n \geq 4$; 9 follicles per mouse were analysed). **c**, Histological analysis and immunostaining for filaggrin and loricrin (red) in 10–12-months-old Bmal1WT and Bmal1KO mice. Graph shows thickness quantification of the cornified layer ($n = 13$). Scale bars, 100 μm. Results are shown as mean ± s.e.m., $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ (two-tailed Student's *t*-test). NS, not significant.

microarray analysis of purified basal interfollicular epidermal progenitors ($\alpha6^{bright}$/CD34^{neg}), from 10-month-old Bmal1KO mice and their control littermates (Supplementary Table 3). As expected, cells from Bmal1KO mice showed strong differential expression of most of the core circadian transcripts, including the microRNA miRNA-122 (ref. 32), some of which were validated independently by real-time quantitative PCR with reverse transcription (RT–qPCR) (Supplementary Fig. 14a and Supplementary Table 3). Interestingly, Gene Ontology (GO) analysis indicated that the cell cycle, energy and drug metabolism, calcium-sensing proteins, the epidermal barrier response and chromatin compaction were significantly affected upon deletion of *Bmal1* (Supplementary Fig. 14b). Intriguingly, although Bmal1KO mice showed a hyperkeratotic phenotype, the viable epidermal layers expressed lower levels of terminal differentiation markers, including *Flg*, *Lor*, *Sprr1*, *Lce* genes and *Tgm* (Supplementary Fig. 14c and Supplementary Table 3). Interestingly, Bmal1KO cells also expressed lower amounts of the miRNA-23b/-27b/-24-1 cluster, which targets TGFβR2 and Smad proteins[33] (Supplementary Fig. 14d). The reduced levels of epidermal differentiation genes probably reflects the lower efficiency of activation of basal interfollicular epidermal cells in Bmal1KO mice, suggesting that the hyperkeratotic phenotype developed as a compensatory mechanism to ensure a certain degree of epidermal barrier protection.

Loss of Bmal1 reduces skin tumorigenesis

Because perturbation of the clock machinery affects the predisposition of certain tissues to carcinogenesis[34], we next studied whether epidermal deletion of *Bmal1* had any impact on the development of cutaneous squamous tumours. To this end, we crossed Bmal1KO mice with a transgenic line expressing oncogenic Sos, an activator of Ras, under the regulation of the Krt5 promoter (K5-SOS)[35]. In an EGFR mutant-heterozygous background, K5-SOS mice spontaneously developed squamous tumours, primarily in the tail, with 100% penetrance, as previously described[35]. Bmal1KO/K5-SOS mice developed significantly fewer neoplastic lesions at early-, mid- and late-stages of carcinoma development than control mice (Fig. 5a and Supplementary Fig. 15a). The skin lesions of Bmal1KO/K5-SOS mice were more differentiated—as determined by increased expression of involucrin and loricrin—contained large cornified islands and a higher number of apoptotic areas, as compared to control tumours (Supplementary Fig. 15b). Control mice had to be killed by two months of age, a time at which no Bmal1KO mice had developed the number, or size, of tumours. The growth of cutaneous squamous tumours has been shown to depend on Wnt activity in a population of CD34$^+$ tumour-initiating cells[36]. However, we could not detect any nuclear β-catenin, either in control or in Bmal1KO neoplastic lesions, suggesting that, in our model, tumour growth did not primarily depend on misregulated Wnt signalling (Supplementary Fig. 16a, b). We did observe a significant reduction in the percentage of $\alpha6^{bright}$/CD34$^+$ tumour-initiating cells in BmalKO tumours with respect to wild-type tumours (Fig. 5b and Supplementary Fig. 17). Notably, whereas wild-type tumours expressed α6 integrin in basal and suprabasal cells, which has been previously associated with increased malignancy[37], Bmal1KO tumours only expressed it in basal

©2011 Macmillan Publishers Limited. All rights reserved

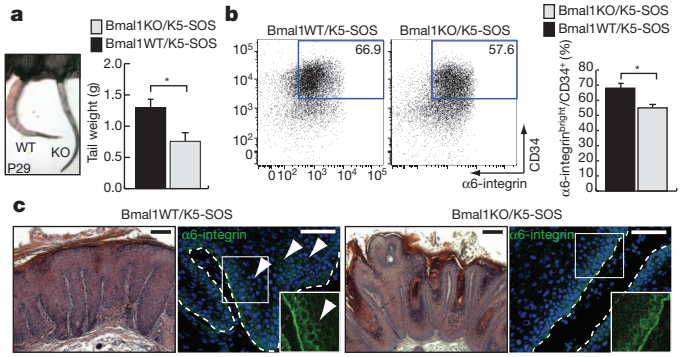


Figure 5 | Loss of Bmal1 reduces the development of squamous tumours. **a**, Quantification of tumour weight of Bmal1WT/K5-SOS and Bmal1KO/K5-SOS littermates ($n \geq 5$). **b**, FACS analysis and quantification of $\alpha6^{bright}/CD34^{+}$ tumour-initiating cells (P36 mice). Numbers represent percentage of $\alpha6^{bright}/CD34^{+}$ cells from gated epithelial cells ($n \geq 3$) **c**, Histological analysis and immunostaining for α6-integrin (arrows indicate suprabasal expression). Scale bars, 50 μm. Results in **a** and **b** are shown as mean ± s.e.m., $^{*}P < 0.05$ (two-tailed Student's *t*-test).

cells (Fig. 5c). The reduction in the percentage of tumour-initiating cells, together with the high expression of several tumour suppressors, in Bmal1KO epidermal progenitors, including the non-coding RNA H19 (top upregulated gene; Supplementary Table 3), probably contribute to the reduced burden of squamous lesions in Bmal1KO mice.

Discussion

Our results indicate that the molecular clock establishes stem cell states that are differentially predisposed to respond to activation and dormancy stimuli. The clock machinery controls the expression of essential epidermal stem cell regulators to anticipate the requirements of the tissue. In this sense, in the epidermal compartment where quiescent stem cells exist (that is, the bulge), the clock establishes a population of 'ready-to-go' cells that can rapidly and efficiently respond to activation stimuli, while simultaneously preventing all stem cells within the niche from becoming responsive. Future studies will be necessary to determine when, and how, this stem cell heterogeneity is established. On the other hand, the murine interfollicular epidermis, which primarily depends on continuously cycling basal progenitors rather than dormant stem cells[38] for its renewal, shows a much more homogenous clock activity (albeit containing approximately 5% of cells antiphasic with the majority). In this compartment the clock machinery might predominantly establish a correct timing of stem cell activation and differentiation. Conditional deletion of *Bmal1* in liver, retina and pancreas results in profound defects in tissue function[39–41], and haematopoietic stem cells show a Bmal1/Clock-dependent circadian release to the periphery[42]. Altogether, these findings indicate that the clock machinery may constitute a fine-tuning homeostatic mechanism in tissues in which dormant and active populations of stem cells coexist. Unbalancing the epidermal stem cell clock not only substantially affected long-term tissue homeostasis, but also the predisposition of the tissue to undergo neoplastic transformation. It is likely that perturbations of this clock-controlled mechanism over stem cell regulation in humans may have long-term consequences on tissue homeostasis, ageing and carcinogenesis.

METHODS SUMMARY

For the isolation of epidermal cells from back or tail skin, the skin was incubated in 0.25% trypsin for 2 h at 37 °C, or overnight at 4 °C, to separate the dermis from the epidermis. Back and tail keratinocytes were extracted as described previously[43].

For ChIP assays, cells in suspension were cross-linked for 10 min at room temperature (22–25 °C) in 1% formaldehyde. Cross-linking reactions were stopped by adding 1.25 M glycine to a final concentration of 125 mM. Cells were centrifuged at 300*g* for 10 min at 4 °C and washed in cold PBS. Cell lysis, sonification and ChIP assays were performed using the MAGnify Chromatin Immunoprecipitation System (Invitrogen).

For FACS analysis or sorting of bulge and epidermal stem cells, cell suspensions were incubated for 30 min on ice with PE-conjugated anti-α6-integrin (CD49f clone NKI-GoH3, Serotec) and biotin-conjugated anti-CD34 (clone RAM34, BD Pharmingen) antibodies followed by APC-conjugated streptavidin (BD Pharmingen) for 20 min. Dead cells were excluded by 4′,6-diamidino-2-phenylindole (DAPI) incorporation. FACS analysis and sorting were performed using LSRII FACS Analysers, FACSAriaII, FACSDiva (BD Biosciences) and Flowjo software.

For microarray analysis, total RNA was isolated from FACS-sorted cells using Trizol extraction and RNeasy Micro Kit (Qiagen). Transcriptional profiling was performed using GeneChip Mouse Gene 1.0 ST Array (Affymetrix) and functional analysis was performed using DAVID Bioinformatics Resources 6.7.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 February; accepted 19 October 2011.
Published online 9 November; corrected 7 December 2011 (see full-text HTML version for details).

1. Fuchs, E. The tortoise and the hair: slow-cycling cells in the stem cell race. *Cell* **137,** 811–819 (2009).
2. Morris, R. J. *et al.* Capturing and profiling adult hair follicle stem cells. *Nature Biotechnol.* **22,** 411–417 (2004).
3. Tumbar, T. *et al.* Defining the epithelial stem cell niche in skin. *Science* **303,** 359–363 (2004).
4. Jaks, V. *et al.* Lgr5 marks cycling, yet long-lived, hair follicle stem cells. *Nature Genet.* **40,** 1291–1299 (2008).
5. Snippert, H. J. *et al.* Lgr6 marks stem cells in the hair follicle that generate all cell lineages of the skin. *Science* **327,** 1385–1389 (2010).
6. Jensen, K. B. *et al.* Lrig1 expression defines a distinct multipotent stem cell population in mammalian epidermis. *Cell Stem Cell* **4,** 427–439 (2009).
7. Jensen, U. B. *et al.* A distinct population of clonogenic and multipotent murine follicular keratinocytes residing in the upper isthmus. *J. Cell Sci.* **121,** 609–617 (2008).
8. Clayton, E. *et al.* A single type of progenitor cell maintains normal epidermis. *Nature* **446,** 185–189 (2007).
9. Kobielak, K., Stokes, N., de la Cruz, J., Polak, L. & Fuchs, E. Loss of a quiescent niche but not follicle stem cells in the absence of bone morphogenetic protein signaling. *Proc. Natl Acad. Sci. USA* **104,** 10063–10068 (2007).
10. Guasch, G. *et al.* Loss of TGFβ signaling destabilizes homeostasis and promotes squamous cell carcinomas in stratified epithelia. *Cancer Cell* **12,** 313–327 (2007).
11. Plikus, M. V. *et al.* Cyclic dermal BMP signalling regulates stem cell activation during hair regeneration. *Nature* **451,** 340–344 (2008).
12. Zhang, Y. V., Cheong, J., Ciapurin, N., McDermitt, D. J. & Tumbar, T. Distinct self-renewal and differentiation phases in the niche of infrequently dividing hair follicle stem cells. *Cell Stem Cell* **5,** 267–278 (2009).
13. Greco, V. *et al.* A two-step mechanism for stem cell activation during hair regeneration. *Cell Stem Cell* **4,** 155–169 (2009).
14. Enshell-Seijffers, D., Lindon, C., Kashiwagi, M. & Morgan, B. A. β-Catenin activity in the dermal papilla regulates morphogenesis and regeneration of hair. *Dev. Cell* **18,** 633–642 (2010).
15. Brownell, I., Guevara, E., Bai, C. B., Loomis, C. A. & Joyner, A. L. Nerve-derived sonic hedgehog defines a niche for hair follicle stem cells capable of becoming epidermal stem cells. *Cell Stem Cell* **8,** 552–565 (2011).
16. Flores, I., Cayuela, M. L. & Blasco, M. A. Effects of telomerase and telomere length on epidermal stem cell behavior. *Science* **309,** 1253–1256 (2005).
17. Owens, D. M. & Watt, F. M. Contribution of stem cells and differentiated cells to epidermal tumours. *Nature Rev. Cancer* **3,** 444–451 (2003).
18. Watt, F. M., Frye, M. & Benitah, S. A. MYC in mammalian epidermis: how can an oncogene stimulate differentiation? *Nature Rev. Cancer* **8,** 234–242 (2008).
19. Benitah, S. A., Frye, M., Glogauer, M. & Watt, F. M. Stem cell depletion through epidermal deletion of Rac1. *Science* **309,** 933–935 (2005).
20. Morris, R. J. A perspective on keratinocyte stem cells as targets for skin carcinogenesis. *Differentiation* **72,** 381–386 (2004).
21. Dibner, C., Schibler, U. & Albrecht, U. The mammalian circadian timing system: organization and coordination of central and peripheral clocks. *Annu. Rev. Physiol.* **72,** 517–549 (2010).
22. Cheng, H. Y. *et al.* Segregation of expression of mPeriod gene homologs in neurons and glia: possible divergent roles of mPeriod1 and mPeriod2 in the brain. *Hum. Mol. Genet.* **18,** 3110–3124 (2009).
23. Kuhlman, S. J., Quintero, J. E. & McMahon, D. G. GFP fluorescence reports period 1 circadian gene regulation in the mammalian biological clock. *Neuroreport* **11,** 1479–1482 (2000).
24. Schneider, M. R., Schmidt-Ullrich, R. & Paus, R. The hair follicle as a dynamic miniorgan. *Curr. Biol.* **19,** R132–R142 (2009).
25. Bunger, M. K. *et al. Mop3* is an essential component of the master circadian pacemaker in mammals. *Cell* **103,** 1009–1017 (2000).
26. Lin, K. K. *et al.* Circadian clock genes contribute to the regulation of hair follicle cycling. *PLoS Genet.* **5,** e1000573 (2009).

©2011 Macmillan Publishers Limited. All rights reserved

27. Kondratov, R. V., Kondratova, A. A., Gorbacheva, V. Y., Vykhovanets, O. V. & Antoch, M. P. Early aging and age-related pathologies in mice deficient in BMAL1, the core componentof the circadian clock. *Genes Dev.* **20,** 1868–1873 (2006).
28. Kameda, T. & Sugiyama, T. Application of genetically modified feeder cells for culture of keratinocytes. *Methods Mol. Biol.* **289,** 29–38 (2005).
29. Zheng, B. *et al.* Nonredundant roles of the *mPer1* and *mPer2* genes in the mammalian circadian clock. *Cell* **105,** 683–694 (2001).
30. Barradas, M. *et al.* Histone demethylase JMJD3 contributes to epigenetic control of *INK4a/ARF* by oncogenic RAS. *Genes Dev.* **23,** 1177–1182 (2009).
31. Ito, M. *et al.* Stem cells in the hair follicle bulge contribute to wound repair but not to homeostasis of the epidermis. *Nature Med.* **11,** 1351–1354 (2005).
32. Gatfield, D. *et al.* Integration of microRNA miR-122 in hepatic circadian gene expression. *Genes Dev.* **23,** 1313–1326 (2009).
33. Rogler, C. E. *et al.* MicroRNA-23b cluster microRNAs regulate transforming growth factor-β/bone morphogenetic protein signaling and liver stem cell differentiation by targeting Smads. *Hepatology* **50,** 575–584 (2009).
34. Fu, L. & Lee, C. C. The circadian clock: pacemaker and tumour suppressor. *Nature Rev. Cancer* **3,** 350–361 (2003).
35. Sibilia, M. *et al.* The EGF receptor provides an essential survival signal for SOS-dependent skin tumor development. *Cell* **102,** 211–220 (2000).
36. Malanchi, I. *et al.* Cutaneous cancer stem cell maintenance is dependent on β-catenin signalling. *Nature* **452,** 650–653 (2008).
37. Owens, D. M., Romero, M. R., Gardner, C. & Watt, F. M. Suprabasal α6β4 integrin expression in epidermis results in enhanced tumourigenesis and disruption of TGFβ signalling. *J. Cell Sci.* **116,** 3783–3791 (2003).
38. Clayton, E. *et al.* A single type of progenitor cell maintains normal epidermis. *Nature* **446,** 185–189 (2007).
39. Storch, K. F. *et al.* Intrinsic circadian clock of the mammalian retina: importance for retinal processing of visual information. *Cell* **130,** 730–741 (2007).
40. Marcheva, B. *et al.* Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes. *Nature* **466,** 627–631 (2010).
41. Lamia, K. A., Storch, K. F. & Weitz, C. J. Physiological significance of a peripheral tissue circadian clock. *Proc. Natl Acad. Sci. USA* **105,** 15172–15177 (2008).
42. Méndez-Ferrer, S., Lucas, D., Battista, M. & Frenette, P. S. Haematopoietic stem cell release is regulated by circadian oscillations. *Nature* **452,** 442–447 (2008).
43. Jensen, K. B., Driskell, R. R. & Watt, F. M. Assaying proliferation and differentiation capacity of stem cells using disaggregated adult mouse epidermis. *Nature Protocols* **5,** 898–911 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the AICR (Association for International Cancer Research), the Spanish Ministry of Health (FIS) and AGAUR (Agència de Gestió d'Ajuts Universitaris i de Recerca; Government of Cataluña) for financial support. P.J. is the recipient of an AGAUR PhD Fellowship, and G.P. of a FIS fellowship. We thank D. McMahon (Vanderbilt University) for providing us with the Per1–GFP mice; E. Wagner (CNIO) for the K5-SOS mice; B. Kübler, the FACS and Genomics units of the IRB (Institute de Recerca Biomedica), the CRG (Center for Genomic Regulation) core facilities and the Animal Unit (Juan Martin Caballero) for technical support.

Author Contributions P.J. performed the experiments, and P.J. and S.A.B. analysed the results and wrote the manuscript. G.P. performed the analysis of K5-SOS mice, and A.M. and E.B. assisted P.J. in the initial FACS sorts. L.D.C. helped us with the initial ChIP experiments. K.O. provided the Per1–venus mice and H.-Y.M.C. generated the Per1–venus mice. J.R. and U.A. provided the Per1/Per2dKO mice.

Author Information Microarray data can be retrieved from the Gene Expression Omnibus under accession number GSE27079. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.A.B. (salvador.aznar-benitah@crg.es).

©2011 Macmillan Publishers Limited. All rights reserved

METHODS

Animals. Bmal1$^{LoxP/LoxP}$ and K14Cre mice purchased from The Jackson Laboratory were crossed with each other to obtain K14Cre/Bmal1$^{wt/wt}$ (Bmal1WT) and K14Cre/Bmal1$^{LoxP/LoxP}$ (Bmal1KO), and with K5-SOS mice to generate K14Cre/Bmal1$^{wt/wt}$/K5-SOS (Bmal1WT/K5SOS) and K14Cre/Bmal1$^{LoxP/LoxP}$/K5-SOS (Bmal1KO/K5SOS) littermate controls[35]. Per1–GFP and Per1–venus mice have been described previously[22,23]. K14Cre/Bmal1$^{wt/wt}$ Per1–venus and K14Cre/Bmal1$^{LoxP/LoxP}$ Per1–venus mice were generated by crossing K14Cre/Bmal1$^{wt/wt}$ and K14Cre/Bmal1$^{LoxP/LoxP}$ animals with Per1–venus mice, respectively. *Per1* and *Per2* double-mutant mice (Per1/2dKO) have been described previously, and were compared to age and sex-matched control animals to avoid differences in hair cycle behaviour between males and females[29]. Mice were housed in an AAALAC-I approved animal unit under 12 h light/12 h dark or 12 h dark/12 h dark cycles, and SPF conditions, and all procedures were approved by the CEEA (Ethical Committee for Animal Experimentation) of the Government of Catalonia. For experiments in constant darkness, light was turned off at Zeitgeber time (ZT)12 and animals were housed in 12 h dark/12 h dark conditions for 5 days. For 5-bromo-2-deoxyuridine (BrdU)-labelling experiments, 100 $\mu g\,g^{-1}$ BrdU (Invitrogen) was injected intraperitoneally into the mice and chased for 10 weeks or up to 10 months, as indicated. To activate epidermal proliferation, back skin or tail skin of 10-week-old Bmal1WT and Bmal1KO mice was treated three times with 20 nM TPA (Sigma-Aldrich) during one week. For depilation experiments, dorsal skin of 15-month-old Bmal1WT and Bmal1KO mice in telogen phase was depilated with cold wax strips (Taky).

Primary mouse keratinocyte cultures. Primary mouse keratinocytes from newborn mice or tail skin of adult mice were isolated as described previously[43,44]. Cells were plated in EMEM (Lonza) containing 4% chelated FBS, 1% penicillin/streptomycin and 20 nM calcium for 24 h; medium was then changed to growth medium (EMEM with 4% chelated FBS, 1% penicillin/streptomycin, EGF (10 ng ml^{-1}) and 50 nM calcium). For the time-course experiments of keratinocytes isolated from Bmal1WT or Bmal1KO mice, cells were synchronized by a serum shock with growth medium, containing 50% chelated FBS, for 2 h[45]. For TGF-β treatments, keratinocytes were cultured in the presence of TGF-β1 (2 ng ml^{-1}; PreproTech) for 48 h. For clonogenic assays, the indicated number of FACS-sorted keratinocytes were plated in E-medium on mitomycin (Sigma-Aldrich)-treated J2 3T3 feeder cells[46]. Keratinocytes were grown at 37 °C with 5% CO_2.

Whole-mount immunofluorescence. Preparation of tail skin and whole-mount stainings were performed as previously described[47]. Primary and secondary antibodies were incubated overnight and used at the following concentrations: 1:1,000 for anti-GFP (A11122, Invitrogen); 1:250 for anti-BrdU (Serotec) and anti-Ki67 (ab15580, Abcam; clone B56, BD Pharmingen); and 1:500 for anti-rabbit and anti-rat conjugated to AlexaFluor488 or AlexaFluor594 (Molecular Probes). Nuclei were stained with DAPI (1:5000; Roche), and epidermal sheets were mounted in Mowiol. Pictures were acquired with a Leica TCS SP5 confocal microscope.

Immunohistochemistry. Back skin and tail skin were either embedded in OCT, or fixed in 4% NBF (Sigma-Aldrich) at 4 °C overnight or 2 h at room temperature (22–25 °C) and then embedded in paraffin or OCT. Deparaffinized sections were boiled for 10 min in 0.01 M citric acid for antigen retrieval. For permeabilization, 8-μm slices of either cryosections, or sections that had been deparaffinized or antigen retrieved were permeabilized for 25 min in 0.25% Triton X-100/PBS and blocked for 90 min in 0.25% gelatin/PBS. Primary antibodies were incubated overnight at 4 °C, and secondary antibodies were incubated for 2 h at room temperature in 0.25% gelatin/PBS. Nuclei were stained with DAPI (1:5,000, Roche), and the slides were mounted in Mowiol. Primary antibodies were used at the following dilutions: 1:50 for anti-p19 (M-167, Santa Cruz), anti-p16 (M-156, Santa Cruz); 1:100 for anti-Sox9 (H-90, Santa Cruz); 1:200 for anti-α6-integrin (CD49f clone NKI-GoH3, Serotec), anti-keratin 14 (ab7800, Abcam); 1:250 for anti-Ki67 (ab15580, Abcam); 1:500 for anti-keratin 5 (ab24647, Abcam); and 1:1,000 for anti-GFP (ab290, Abcam; A11122, Invitrogen), anti-involucrin (PRB-140C, Covance), anti-filaggrin (ab24584, Abcam), anti-loricrin (PRB-145P, Covance), anti-phospho-Smad2 (#3101, Cell Signaling), anti-Lef1 (clone C12A5, Cell Signaling). Secondary antibodies were used at dilutions of 1:500: anti-rabbit, anti-mouse, or anti-rat, conjugated to AlexaFluor488 or AlexaFluor594 (Molecular Probes), or anti-rabbit biotin (GE Healthcare), followed by an incubation with ABC Kit (Vector Laboratories) and FAST 3,3 Diaminobenzidine tablets (Sigma-Aldrich). For apoptosis measurements, tail skin was stained with the DeadEnd Fluorometric TUNEL System (Promega). Staining with haematoxilin or eosin was done according to a standard protocol. Pictures were acquired with a Leica DMI 6000B or a Leica TCS SP5 confocal microscope.

Time-lapse microscopy and quantification. For time-lapse confocal imaging, back skin of Per1–venus mice was fixed with 0.5% agarose in an imaging dish (Ibidi) and over-layered with E-medium. Images were taken every 15 min for a period of 48 h using a Leica TCS SP5 confocal microscope equipped with a tempered chamber of 37 °C and 5% CO_2. Mean fluorescence intensity of individual GFP-positive nuclei was quantified using ImageJ software.

FACS. Epidermal cells from back skin of Per1–venus, Bmal1WT and Bmal1KO mice, or tail skin of Bmal1WT/K5-SOS and Bmal1KO/K5-SOS mice, were isolated as described previously[43]. Per1–venus mice were killed between ZT10–ZT12 or as specified in the figure legends. Cell suspensions were incubated for 30 min on ice with the following antibodies at the given dilutions: 1:100 for biotin- or APC-conjugated anti-CD34 (clone RAM34, BD Pharmingen), PE-conjugated anti-CD31 (clone MEC13.3, eBioscience), PE-conjugated anti-CD45 (clone 30F11, eBioscience), PE-conjugated anti-CD140a (clone APA5, BD Pharmingen) and APC-Cy7-conjugated anti-epcam (clone G8.8, Biolegend); 1:200 for PE- or FITC-conjugated anti-α6-integrin (CD49f clone NKI-GoH3, Serotec); and 1:500 for APC-conjugated streptavidin (BD Pharmingen). DAPI staining was used to rule out that dead cells were present. FACS was performed using FACSAriaII and FACSDiva software (BD Biosciences). Sorted cells were collected in E-medium, either plated in culture or re-suspended in Trizol (Invitrogen), and subjected further to RNA isolation. FACS analysis was performed using LSRII FACS Analysers (BD Biosciences) and Flowjo software.

Arrays. Total RNA was isolated from FACS-sorted cells in Trizol by chloroform extraction, followed by the RNA clean-up protocol of RNeasy Micro or Mini Kit (Qiagen). Transcriptional profiling was performed using GeneChip Mouse Gene 1.0 ST Array (Affymetrix). Arrays of venus^{bright} and venus^{dim} bulge cells of P19 Per1–venus mice were performed as triplicates from a pool of $n = 64$ mice. Arrays of epidermal cells from 10-months-old Bmal1WT and Bmal1KO mice were performed as triplicates using three independent mice in each group. Functional analysis of microarray data was performed using DAVID Bioinformatics Resources 6.7.

RT–qPCR. Total RNA from cultured or FACS-sorted cells was purified as described above, or using either the RNeasy Micro and Mini Kit (Qiagen) or miRVana miRNA isolation Kit (Ambion). Equal amounts of RNA were reverse-transcribed using Superscript III (Invitrogen). RT–qPCR was performed with SYBR Green Master Mix (Roche) and gene-specific primers (as given in Supplementary Table 4)[48] using a Light Cycler 480 Instrument (Roche). Relative levels of expression were determined by normalization to pumilio 1 (Pum1), using the $\Delta\Delta C_t$ method.

ChIP. For ChIP assays from intact epidermis, mice were killed at ZT2, or for time-course analysis every 3 h during a period of 24 h, and tails were incubated in 0.25% trypsin for 2 h at 37 °C to separate the dermis from the epidermis. Tail keratinocytes were extracted as described[43]. Cells in suspension or FACS-purified cells (2×10^5) were cross-linked for 10 min at room temperature in 1% formaldehyde. Cross-linking reactions were stopped by adding 1.25 M glycine to a final concentration of 125 mM. Cells were centrifuged at 300*g* for 10 min at 4 °C and washed in cold PBS. Cell lysis, sonification and ChIP assays were performed using the MAGnify Chromatin Immunoprecipitation System (Invitrogen). For each immunoprecipitation, 2×10^5–10^6 cells were incubated with 2 μl of anti-Bmal1 or anti-Clock antibody (provided by J. Ripperger, University of Fribourg)[49], or 2 μl of rabbit IgG control antibody (Invitrogen). RT–qPCR was performed as described above using gene-specific primers (Supplementary Table 5).

Genotyping and western blots. Genomic DNA of primary mouse keratinocytes from Bmal1WT and Bmal1KO mice was isolated using a standard protocol. Multiplex PCR was performed as previously described to determine recombination efficiency[39]. Protein extracts from newborn keratinocytes or from tail epidermis were analysed by SDS–PAGE and western blotting for Bmal1 (ref. 49), anti-phospho-Smad2 (1:1,000, #3101, Cell Signaling), anti-Lef1 (1:1,000, clone C12A5, Cell Signaling), anti-Sox9 (1:400, H-90, Santa Cruz) and anti-tubulin (1:5,000, Sigma Aldrich)

Luciferase assay. Primary mouse keratinocytes were transiently transfected with either FOP-Flash or TOP-Flash and pCMV-Renilla plasmids using FuGENE6 (Roche) according to the manufacturer's instructions. The TOP-Flash construct reports Wnt activity by driving the expression of Luciferase under the regulation of TCF-Lef binding sites. FOP-Flash contains the same binding sites mutated so that they are no longer responsive to TCF/Lef. Cells were treated in the absence or presence of 1 μM BIO (Calbiochem) at 48 h post-transfection for a period of 24 h. Luciferase activity was measured in a Centro LB 960 luminometer (Berthold Technologies) using Dual-Luciferase Reporter Assay System (Promega).

Promoter analysis. Gene promoter analysis for potential Bmal1/Clock-binding sites (in general, from −5,000 to +1,000 bases from the transcriptional start site) were analysed using Genomatix Software.

Statistics. Results are presented as mean ± s.e.m. Statistical significance was determined by two-tailed Student's *t*-test, one-way ANOVA, two-way ANOVA

©2011 Macmillan Publishers Limited. All rights reserved

with Bonferroni post-test, or Cosinor analysis. A P value of $P \leq 0.05$ was considered to be statistically significant.

44. Litchi, U., Anders, J. & Yuspa, S. Isolation and short-term culture of primary keratinocytes, hair follicle populations and dermal cells from newborn mice and keratinocytes from adult mice for *in vitro* analysis and for grafting to immunodeficient mice. *Nature Protocols* **5,** 799–810 (2008).
45. Balsalobre, A., Damiola, F. & Schibler, U. A serum shock induces circadian gene expression in mammalian tissue culture cells. *Cell* **93,** 929–937 (1998).
46. Nowak, J. A. & Fuchs, E. Isolation and culture of epithelial stem cells. *Methods Mol. Biol.* **482,** 215–232 (2009).
47. Braun, K. M. *et al.* Manipulation of stem cell proliferation and lineage commitment: visualization of label-retaining cells in wholemounts of mouse epidermis. *Development* **130,** 5241–5255 (2003).
48. Sun, F. *et al.* Characterization of function and regulation of miR-24-1 and miR-31. *Biochem. Biophys. Res. Commun.* **380,** 660–665 (2009).
49. Ripperger, J. & Schibler, U. Rhythmic CLOCK-BMAL1 binding to multiple E-box motifs drives circadian *Dbp* transcription and chromatin transitions. *Nature Genet.* **38,** 369–374 (2006).

©2011 **Macmillan Publishers Limited. All rights reserved**

LETTER

doi:10.1038/nature10636

Two ten-billion-solar-mass black holes at the centres of giant elliptical galaxies

Nicholas J. McConnell[1], Chung-Pei Ma[1], Karl Gebhardt[2], Shelley A. Wright[1], Jeremy D. Murphy[2], Tod R. Lauer[3], James R. Graham[1,4] & Douglas O. Richstone[5]

Observational work conducted over the past few decades indicates that all massive galaxies have supermassive black holes at their centres. Although the luminosities and brightness fluctuations of quasars in the early Universe suggest that some were powered by black holes with masses greater than 10 billion solar masses[1,2], the remnants of these objects have not been found in the nearby Universe. The giant elliptical galaxy Messier 87 hosts the hitherto most massive known black hole, which has a mass of 6.3 billion solar masses[3,4]. Here we report that NGC 3842, the brightest galaxy in a cluster at a distance from Earth of 98 megaparsecs, has a central black hole with a mass of 9.7 billion solar masses, and that a black hole of comparable or greater mass is present in NGC 4889, the brightest galaxy in the Coma cluster (at a distance of 103 megaparsecs). These two black holes are significantly more massive than predicted by linearly extrapolating the widely used correlations between black-hole mass and the stellar velocity dispersion or bulge luminosity of the host galaxy[5–9]. Although these correlations remain useful for predicting black-hole masses in less massive elliptical galaxies, our measurements suggest that different evolutionary processes influence the growth of the largest galaxies and their black holes.

Empirical scaling relations between black-hole mass (M_{BH}), galaxy bulge velocity dispersion (σ) and luminosity (L) are commonly used to estimate black-hole masses, because for most galaxies we are unable to make a direct measurement. Estimates of the number density of black holes in a given mass range thus depend upon the empirically determined M_{BH}–σ and M_{BH}–L relations over an appropriate range of galaxy masses. Directly measuring M_{BH} from the kinematics of stars or gas in the vicinity of the black hole is particularly difficult at the highest galaxy masses, because massive galaxies are rare, their typical distances from Earth are large and their central stellar densities are relatively low. The most massive galaxies are typically brightest cluster galaxies (BCGs), that is, giant elliptical galaxies that reside near the centres of galaxy clusters.

We have obtained high-resolution, two-dimensional data of the line-of-sight stellar velocities in the central regions of NGC 3842 and NGC 4889 using integral-field spectrographs at the Gemini North and Keck 2 telescopes, in Hawaii. The stellar luminosity distribution of each galaxy is provided by surface photometry from NASA's Hubble Space Telescope and ground-based telescopes[10,11]. NGC 3842 is the BCG of Abell 1367, a moderately rich galaxy cluster. NGC 4889 is the BCG of the Coma cluster (Abell 1656), one of the richest nearby galaxy clusters. We targeted these two galaxies because they have relatively high central surface brightnesses and lie at an accessible distance for direct measurements of M_{BH}.

We measured the distribution of stellar velocities at 82 different locations in NGC 3842. The line-of-sight velocity dispersion in NGC 3842 is between 270 and 300 km s^{-1} at large galactocentric radii (r) and rises in the central 0.7 arcsec ($r < 330$ pc), peaking at 326 km s^{-1} (Figs 1 and 2). We determined the mass of the central black hole by constructing a series of orbit superposition models[12]. Each model assumes a black-hole mass, stellar mass-to-light ratio and dark-matter profile, and generates a library of time-averaged stellar orbits in the resulting gravitational potential. The model then fits a weighted combination of orbital line-of-sight velocities to the set of measured stellar velocity distributions. The goodness-of-fit statistic χ^2 is computed as a function of the assumed values of M_{BH} and the stellar mass-to-light ratio. Using our best-fitting model dark-matter halo, we measure a black-hole mass of 9.7×10^9 solar masses ($M_\odot$), with a 68% confidence interval of $(7.2–12.7) \times 10^9 M_\odot$. Models with no black hole are ruled out at the 99.996% confidence level ($\Delta\chi^2 = 17.1$). We find the stellar mass-to-light ratio to equal $5.1 M_\odot/L_\odot$ in the R band ($L_\odot$, solar luminosity), with a 68% confidence interval of 4.4–$5.8 M_\odot/L_\odot$.

We measured stellar velocity distributions at 63 locations in NGC 4889 and combined our measurements with published long-slit kinematics at larger radii[13]. The largest velocity dispersions in NGC 4889 are located across an extended region on the east side of the galaxy. The stellar orbits in our models are defined to be symmetric about the galaxy centre, so we constrain M_{BH} by running separate trials with velocity profiles from four quadrants of the galaxy. The best-fitting black-hole masses from the four quadrants range from $9.8 \times 10^9 M_\odot$ to $2.7 \times 10^{10} M_\odot$. All quadrants favour tangential orbits near the galaxy centre, which cause the line-of-sight velocity dispersion to decrease even as the internal three-dimensional velocity dispersion increases towards the black hole. Although no single model is consistent with all of the observed kinematic features in NGC 4889, we can define a confidence interval for M_{BH} by considering the most extreme confidence limits from the cumulative set of models. The corresponding 68% confidence interval is $(0.6–3.7) \times 10^{10} M_\odot$. We adopt a black-hole mass of $2.1 \times 10^{10} M_\odot$, corresponding to the midpoint of this interval.

Figure 3 shows the M_{BH}–σ and M_{BH}–L relations, using data compiled from studies published before the end of August 2011, plus our measurements of NGC 3842 and NGC 4889. Tabulated data with references are provided as Supplementary Information. The most widely used form for both relations is a power law with a constant exponent. Straight lines in Fig. 3 show our fits to $M_{BH}(\sigma)$ and $M_{BH}(L)$. The relationship between σ and L, however, flattens at high galaxy masses, and constant-exponent power laws for the M_{BH}–σ and M_{BH}–L relations produce contradictory predictions for M_{BH} in this mass range[14]. Direct measurements of M_{BH} in higher mass galaxies will compel the revision of one or both of the M_{BH}–σ and M_{BH}–L relations.

The average velocity dispersion in NGC 3842 is 270 km s^{-1}, measured outside the black hole radius of influence (1.2 arcsec or 570 pc) and inside the two-dimensional half-light radius (38 arcsec or 18 kpc). Although NGC 3842 hosts a black hole more massive than any previously detected, its average dispersion ranks only fourteenth

[1]Department of Astronomy, University of California, Berkeley, California 94720, USA. [2]Department of Astronomy, University of Texas, Austin, Texas 78712, USA. [3]National Optical Astronomy Observatory, Tucson, Arizona 85726, USA. [4]Dunlap Institute for Astronomy and Astrophysics, University of Toronto, Ontario M5S 3H4, Canada. [5]Department of Astronomy, University of Michigan, Ann Arbor, Michigan 48109, USA.

©2011 Macmillan Publishers Limited. All rights reserved

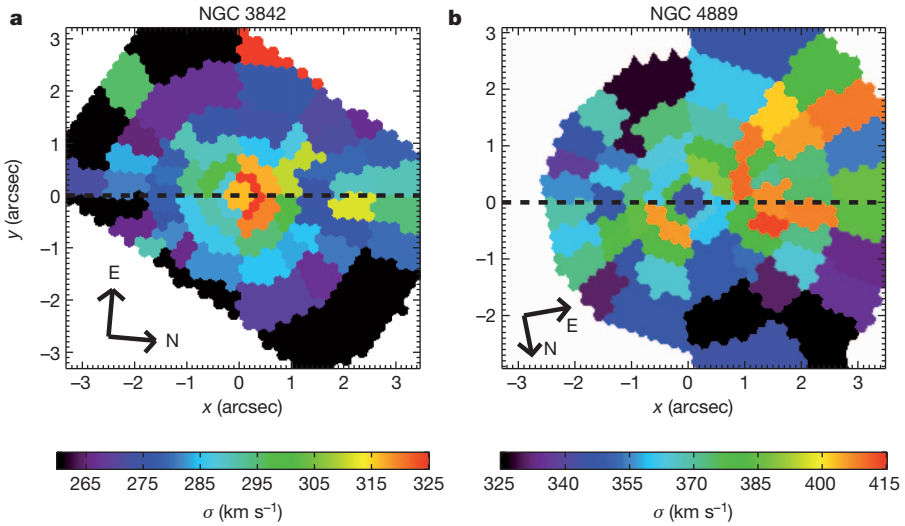
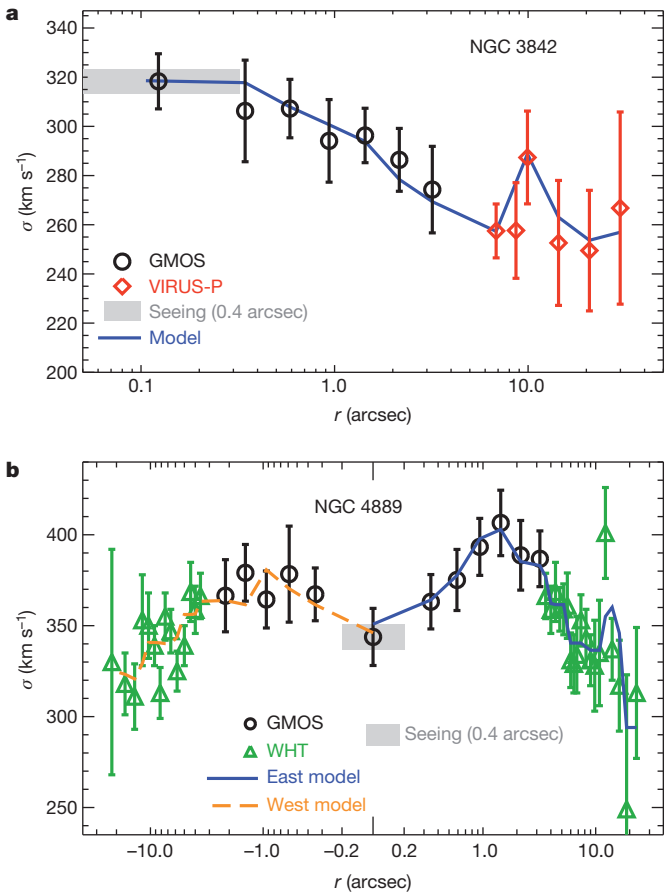


Figure 1 | Two-dimensional maps of the line-of-sight stellar velocity dispersions. The maps show the central regions of NGC 3842 (**a**) and NGC 4889 (**b**) observed using the GMOS spectrograph[23] on the 8-m Gemini North telescope. Additional kinematics at large radii were measured using the VIRUS-P spectrograph[24] on the 2.7-m Harlan J. Smith telescope, and additional high-resolution data were acquired with OSIRIS spectrograph[25] at the 10-m Keck 2 telescope. GMOS, OSIRIS and VIRUS-P are all integral-field spectrographs, which record spectra at multiple positions in a two-dimensional spatial array. The horizontal dashed line in each panel traces the major axis of the galaxy. The median 1-s.d. errors in velocity dispersion are 12 km s^{-1} and 20 km s^{-1} for NGC 3842 and NGC 4889, respectively. In NGC 4889, the highest velocity dispersions, near 410 km s^{-1}, are located on the east side of the galaxy, at least 1.1 arcsec from the centre.

among 65 galaxies with direct measurements of M_{BH}. Its luminosity ranks fifth in this sample of galaxies and is exceeded only by other BCGs. On the basis of the values of σ and L for NGC 3842, our revised M_{BH}–σ and M_{BH}–L relations predict that $M_{BH} = 9.1 \times 10^8 M_\odot$ and $2.5 \times 10^9 M_\odot$, respectively. Similarly, for NGC 4889 the respective predictions are $M_{BH} = 3.3 \times 10^9 M_\odot$ and $4.5 \times 10^9 M_\odot$. These predictions are smaller than our direct measurements of M_{BH} by 1.6–4.6 times the 1-s.d. scatter in the M_{BH}–σ and M_{BH}–L relations[9]. Four measurements of M_{BH} in BCGs existed before this work. Two measurements based on gas dynamics and one based on stellar dynamics all lie within 1.2 s.d. of our revised fits to the M_{BH}–σ and M_{BH}–L relations[15,16]. Yet the measurement of M_{BH} in NGC 1316, the BCG of the Fornax cluster, is 3.4 s.d. less than that predicted by our M_{BH}–L relation[17]. The high scatter indicated by this collection of measurements reveals large uncertainties in the standard practice of using galactic σ or L as a proxy for the central black-hole mass in giant elliptical galaxies and their predecessors.



Several BCGs within 200 Mpc of Earth are at least twice as luminous as NGC 3842 and three times as luminous as Messier 87, which hosted the most massive black hole known before this work. In spite of their extreme luminosities, BCGs have velocity dispersions similar to those of the most massive field elliptical galaxies. Yet the most massive black holes are found predominantly in BCGs (Fig. 3). How galaxies are assembled and the role of gas dissipation affect the correlations (or lack thereof) among M_{BH}, σ and L. Simulations of mergers of gas-rich disk galaxies are able to produce remnant galaxies that follow the observed M_{BH}–σ correlation in Fig. 3a over the intermediate mass range $M_{BH} \approx 10^7 M_\odot$–$10^9 M_\odot$ (refs 18, 19). By contrast, simulated mergers of elliptical galaxies with low-angular-momentum progenitor orbits increase M_{BH} and L by similar numerical factors, without increasing the velocity dispersion[20]. Because these mergers are a likely path to forming the most massive galaxies, the M_{BH}–σ correlation may steepen or disappear altogether at the highest galaxy masses. Massive elliptical galaxies retain residual quantities of gas even after the decline of star formation. Accretion of this gas onto the galaxies' central black

Figure 2 | One-dimensional profiles of the line-of-sight velocity dispersions. **a**, Dispersion versus radius in NGC 3842, after averaging data at a given radius, based on measurements with GMOS (black circles) and VIRUS-P (red diamonds). The solid blue line is the projected line-of-sight dispersion from our best-fitting stellar orbit model of NGC 3842. **b**, Dispersion versus radius along the major axis of NGC 4889, measured using GMOS (black circles) and the William Herschel Telescope[13] (WHT; green triangles). The maximum velocity dispersion occurs at $r = 1.4$ arcsec. The solid blue line is the projected line-of-sight dispersion from our best-fitting orbit model using data from the east side of NGC 4889 ($r > 0$). The dashed orange line is from our best-fitting orbit model using data from the west side of NGC 4889 ($r < 0$). Error bars, 1 s.d.

©2011 Macmillan Publishers Limited. All rights reserved

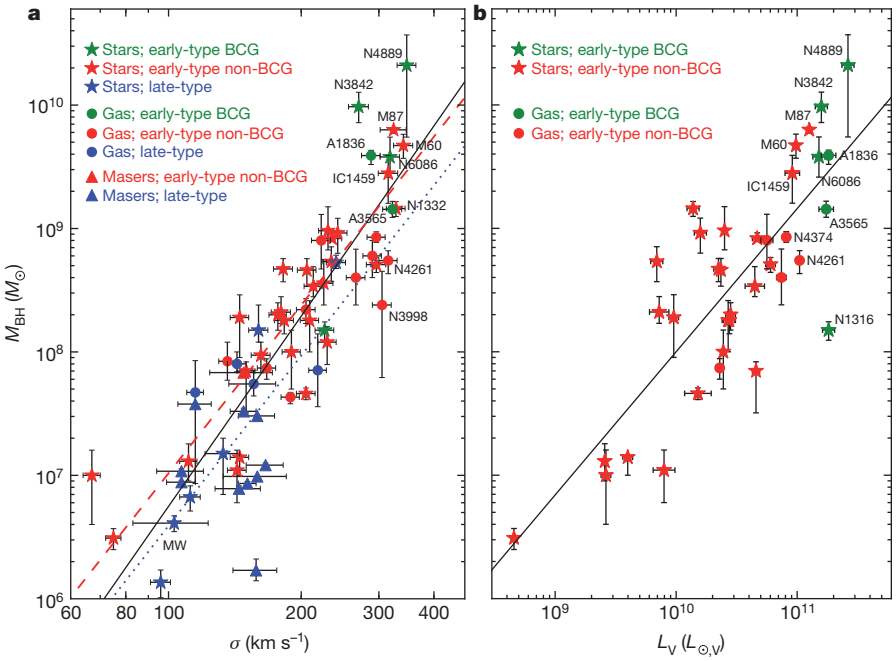


Figure 3 | Correlations of dynamically measured black hole masses and bulk properties of host galaxies. **a**, Black-hole mass (M_{BH}) versus stellar velocity dispersion (σ) for 65 galaxies with direct dynamical measurements of M_{BH}. For galaxies with spatially resolved stellar kinematics, σ is the luminosity-weighted average within one effective radius (Supplementary Information). **b**, Black-hole mass versus V-band bulge luminosity, L_V ($L_{\odot,V}$, solar value), for 36 early-type galaxies with direct dynamical measurements of M_{BH}. Our sample of 65 galaxies consists of 32 measurements from a 2009 compilation[9], 16 galaxies with masses updated since 2009, 15 new galaxies with M_{BH} measurements and the two galaxies reported here. A complete list of the galaxies is given in Supplementary Table 4. BCGs (defined here as the most luminous galaxy in a cluster) are plotted in green, other elliptical and S0 galaxies are plotted in red, and late-type spiral galaxies are plotted in blue. The black-hole masses are measured using the dynamics of masers (triangles), stars (stars) or gas (circles). Error bars, 68% confidence intervals. For most of the maser galaxies, the error bars in M_{BH} are smaller than the plotted symbol. The solid black line in **a** shows the best-fitting power law for the entire sample: $\log_{10}(M_{BH}/M_\odot) = 8.29 + 5.12\log_{10}[\sigma/(200\,\mathrm{km\,s^{-1}})]$. When early-type and late-type galaxies are fitted separately, the resulting power laws are $\log_{10}(M_{BH}/M_\odot) = 8.38 + 4.53\log_{10}[\sigma/(200\,\mathrm{km\,s^{-1}})]$ for elliptical and S0 galaxies (dashed red line) and $\log_{10}(M_{BH}/M_\odot) = 7.97 + 4.58\log_{10}[\sigma/(200\,\mathrm{km\,s^{-1}})]$ for spiral galaxies (dotted blue line). The solid black line in **b** shows the best-fitting power law: $\log_{10}(M_{BH}/M_\odot) = 9.16 + 1.16\log_{10}(L_V/10^{11}L_\odot)$. We do not label Messier 87 as a BCG, as is commonly done, because NGC 4472 in the Virgo cluster is 0.2 mag brighter.

holes could help increase M_{BH} and further steepen the M_{BH}–σ and M_{BH}–L relations.

Black holes in excess of $10^{10}M_\odot$ are observed as quasars in the early Universe, from 1.4×10^9 to 3.3×10^9 yr after the Big Bang[2] (redshift, $z = 2$–4.5). Throughout the last 1.0×10^{10} yr, however, these extremely massive black holes have not been accreting appreciably, and the average mass of the black holes powering quasars has decreased steadily. Quasar activity and elliptical galaxy formation are predicted to arise from similar merger-triggered processes, and there is growing evidence that present-day massive elliptical galaxies once hosted the most-luminous high-redshift quasars[21]. However, definitive classification of these quasars' host galaxies has remained elusive.

Our measurements of black holes with masses of around $10^{10}M_\odot$ in NGC 3842 and NGC 4889 provide circumstantial evidence that BCGs host the remnants of extremely luminous quasars. The number density of nearby BCGs ($\sim5\times10^{-6}\,\mathrm{Mpc^{-3}}$) is consistent with the number density of black holes ($\sim3\times10^{-7}$ to $10^{-5}\,\mathrm{Mpc^{-3}}$) with masses between $10^9M_\odot$ and $10^{10}M_\odot$ predicted from the M_{BH}–L relation and the luminosity function of nearby galaxies. Furthermore, both quantities agree with predictions based on the black-hole masses and duty cycles of quasars. The black-hole number density predicted from the M_{BH}–σ relation, however, is an order of magnitude less than the inferred quasar population[14,22]. These two predictions can be reconciled if the M_{BH}–σ relation has upward curvature or a large degree of intrinsic scatter in M_{BH} at the high-mass end, as suggested by our new measurements. With improvements in adaptive optics instrumentation on large optical telescopes and very-long-baseline interferometry at radio wavelengths, black holes are being sought and detected in increasingly exotic host galaxies. Along with our measurements of the black-hole masses in NGC 3842 and NGC 4889, future measurements in other massive galaxies will quantify the cumulative growth of supermassive black holes in the Universe's densest environments.

Received 15 July; accepted 13 October 2011.

1. Netzer, H. The largest black holes and the most luminous galaxies. *Astrophys. J.* **583,** L5–L8 (2003).
2. Vestergaard, H., Fan, X., Tremonti, C. A., Osmer, P. S. & Richards, G. T. Mass functions of the active black holes in distant quasars from the Sloan Digital Sky Survey data release 3. *Astrophys. J.* **674,** L1–L4 (2008).
3. Sargent, W. L. W. *et al.* Dynamical evidence for a central mass concentration in the galaxy M87. *Astrophys. J.* **221,** 731–744 (1978).
4. Gebhardt, K. *et al.* The black hole mass in M87 from Gemini/NIFS adaptive optics observations. *Astrophys. J.* **729,** 119 (2011).
5. Dressler, A. in *Active Galactic Nuclei* (eds Osterbrock, D. E. & Miller, J. S.) 217–232 (IAU Symposium 134, Springer, 1989).
6. Kormendy, J. & Richstone, D. Inward bound the search for supermassive black holes in galactic nuclei. *Annu. Rev. Astron. Astrophys.* **33,** 581–624 (1995).
7. Ferrarese, L. & Merritt, D. A fundamental relation between supermassive black holes and their host galaxies. *Astrophys. J.* **539,** L9–L12 (2000).
8. Gebhardt, K. *et al.* A relationship between nuclear black hole mass and galaxy velocity dispersion. *Astrophys. J.* **539,** L13–L16 (2000).
9. Gültekin, K. *et al.* The M_{BH}–σ and M_{BH}–L relations in galactic bulges, and determinations of their intrinsic scatter. *Astrophys. J.* **698,** 198–221 (2009).
10. Laine, S. *et al.* Hubble Space Telescope imaging of brightest cluster galaxies. *Astron. J.* **125,** 478–505 (2003).
11. Postman, M. & Lauer, T. R. Brightest cluster galaxies as standard candles. *Astrophys. J.* **440,** 28–47 (1995).
12. Schwarzschild, M. A numerical model for a triaxial stellar system in dynamical equilibrium. *Astrophys. J.* **232,** 236–247 (1979).
13. Loubser, S. I., Sansom, A. E., Sánchez-Blázquez, P., Soechting, I. K. & Bromage, G. E. Radial kinematics of brightest cluster galaxies. *Mon. Not. R. Astron. Soc.* **391,** 1009–1028 (2008).
14. Lauer, T. R. *et al.* The masses of nuclear black holes in luminous elliptical galaxies and implications for the space density of the most massive black holes. *Astrophys. J.* **662,** 808–834 (2007).

©2011 Macmillan Publishers Limited. All rights reserved

15. Dalla Bontà, E. *et al.* The high-mass end of the black hole mass function: mass estimates in brightest cluster galaxies. *Astrophys. J.* **690,** 537–559 (2009).
16. McConnell, N. J. *et al.* The black hole mass in brightest cluster galaxy NGC 6086. *Astrophys. J.* **728,** 100 (2011).
17. Nowak, N. *et al.* The supermassive black hole of Fornax A. *Mon. Not. R. Astron. Soc.* **391,** 1629–1649 (2008).
18. Di Matteo, T., Volker, S. & Hernquist, L. Energy input from quasars regulates the growth and activity of black holes and their host galaxies. *Nature* **433,** 604–607 (2005).
19. Robertson, B. *et al.* The evolution of the $M_{BH} - \sigma$ relation. *Astrophys. J.* **641,** 90–102 (2006).
20. Boylan-Kolchin, M., Ma, C.-P. & Quataert, E. Red mergers and the assembly of massive elliptical galaxies: the fundamental plane and its projections. *Mon. Not. R. Astron. Soc.* **369,** 1081–1089 (2006).
21. Hopkins, P. F., Bundy, K., Hernquist, L. & Ellis, R. S. Observational evidence for the coevolution of galaxy mergers, quasars, and the blue/red galaxy transition. *Astrophys. J.* **659,** 976–996 (2007).
22. Lauer, T. R., Tremaine, S., Richstone, D. & Faber, S. M. Selection bias in observing the cosmological evolution of the $M_{BH}-\sigma$ and $M_{BH}-L$ relationships. *Astrophys. J.* **670,** 249–260 (2007).
23. Allington-Smith, J. *et al.* Integral field spectroscopy with the Gemini Multi-Object Spectrograph. I. Design, construction, and testing. *Publ. Astron. Soc. Pacif.* **114,** 892–912 (2002).
24. Hill, G. J. *et al.* Design, construction, and performance of VIRUS-P: the prototype of a highly replicated integral field spectrograph for the HET. *Proc. Soc. Photo-opt. Instrum. Eng.* **7014,** 701470 (2008).
25. Larkin, J. *et al.* OSIRIS: a diffraction limited integral field spectrograph for Keck. *Proc. Soc. Photo-opt. Instrum. Eng.* **6269,** 62691A (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements N.J.M., C.-P.M, K.G. and J.R.G. are supported by the National Science Foundation. C.-P.M. is supported by NASA and by the Miller Institute for Basic Research in Science, University of California, Berkeley. S.A.W. is supported by NASA through a Hubble Fellowship. Data presented here were obtained using the Gemini Observatory, the W. M. Keck Observatory and the McDonald Observatory. The Gemini Observatory is operated by the Association of Universities for Research in Astronomy, Inc., under a cooperative agreement with the National Science Foundation on behalf of the Gemini partnership. The W. M. Keck Observatory is operated as a scientific partnership among the California Institute of Technology, the University of California and NASA. The McDonald Observatory is operated by the University of Texas at Austin. The instrument VIRUS-P was funded by G. and C. Mitchell. Stellar orbit models were run using the facilities at the Texas Advanced Computing Center at The University of Texas at Austin.

Author Contributions N.J.M. carried out the data analysis and modelling. N.J.M, C.-P.M. and S.A.W. wrote the manuscript. C.-P.M. compiled the data for Fig. 3 and oversaw communication among co-authors. S.A.W. analysed OSIRIS data on NGC 3842. K.G. provided GMOS data on NGC 3842 and NGC 4889. K.G. and D.O.R. developed the stellar orbit modelling code. J.D.M. provided VIRUS-P data on NGC 3842. T.R.L. provided photometric data on and image analysis of NGC 3842 and NGC 4889. J.R.G. led the OSIRIS observing campaign for NGC 3842. All authors contributed to the interpretive analysis of the observations and the writing of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.J.M. (nmcc@berkeley.edu) or C.-P.M. (cpma@berkeley.edu).

©2011 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature10654

Atomic homodyne detection of continuous-variable entangled twin-atom states

C. Gross[1], H. Strobel[1], E. Nicklas[1], T. Zibold[1], N. Bar-Gill[2]†, G. Kurizki[2] & M. K. Oberthaler[1]

Historically, the completeness of quantum theory has been questioned using the concept of bipartite continuous-variable entanglement[1]. The non-classical correlations (entanglement) between the two subsystems imply that the observables of one subsystem are determined by the measurement choice on the other, regardless of the distance between the subsystems. Nowadays, continuous-variable entanglement is regarded as an essential resource, allowing for quantum enhanced measurement resolution[2], the realization of quantum teleportation[3–5] and quantum memories[3,6], or the demonstration of the Einstein–Podolsky–Rosen paradox[1,7–9]. These applications rely on techniques to manipulate and detect coherences of quantum fields, the quadratures. Whereas in optics coherent homodyne detection[10] of quadratures is a standard technique, for massive particles a corresponding method was missing. Here we report the realization of an atomic analogue to homodyne detection for the measurement of matter-wave quadratures. The application of this technique to a quantum state produced by spin-changing collisions in a Bose–Einstein condensate[11,12] reveals continuous-variable entanglement, as well as the twin-atom character of the state[13]. Our results provide a rare example of continuous-variable entanglement of massive particles[6,14]. The direct detection of atomic quadratures has applications not only in experimental quantum atom optics, but also for the measurement of fields in many-body systems of massive particles[15].

Continuous-variable entangled states, whose inter-mode entanglement is reflected in quadrature as well as population correlations, are routinely generated by parametric down-conversion in optical parametric amplifiers (OPAs)[10]. Spin-changing collisions in ultracold bosonic quantum gases provide a similar nonlinear process for matter waves, generating atom pairs with one atom spin-up and the other spin-down—that is, twin atoms[16–19]. Correlated photon or atom pairs in the signal $|\uparrow\rangle$ and idler $|\downarrow\rangle$ modes are created from a pump field via nonlinear interactions described by the Hamiltonian $\hbar\alpha_0\left(a_\downarrow^\dagger a_\uparrow^\dagger + a_\downarrow a_\uparrow\right)$, where $a_k^\dagger$ (a_k) is the creation (annihilation) operator of the respective mode k and $2\pi\hbar$ is Planck's constant. This Hamiltonian presumes a large-amplitude coherent pump field, whose depletion or distortion by the signal and idler modes is negligible, such that it can be treated classically. In optics, the effective nonlinearity α_0 is proportional to the amplitude of this pump field and the susceptibility of the medium. In the analogous regime for atoms, the effective nonlinearity originates from interatomic interactions and is proportional to the population in the pump mode, that is, the Bose–Einstein condensate (BEC). For initially empty signal and idler modes, quantum fluctuations are amplified and the output is the so-called two-mode squeezed vacuum[10]. This state is characterized by the vanishing of the mean field amplitude $\langle E_k\rangle = 0$ in each of the modes, while the mean intensity is non-zero $\langle I_k\rangle \propto \left\langle E_k^\dagger E_k\right\rangle > 0$. The associated two-mode entanglement is revealed by the quadratures $X_\pm(\varphi) = a_\downarrow e^{-i\varphi} \pm a_\uparrow^\dagger e^{i\varphi} + \text{h.c.}$, the signature being the squeezing of their variance for appropriately chosen phase φ (h.c. indicates the Hermitian conjugate terms). To access the two-mode quadratures, the coherent homodyne measurement technique is employed, whereby the conjugate canonical quadratures $X_k = a_k^\dagger + a_k$ and $Y_k = i\left(a_k^\dagger - a_k\right)$ are measured by mixing the quantum mode with a strong classical reference field, the local oscillator $a_{\mathrm{LO}}^\dagger \approx \sqrt{N_{\mathrm{LO}}}e^{-i\varphi}$ with large amplitude $\sqrt{N_{\mathrm{LO}}}$. Experimental control of the local oscillator phase φ provides access to the continuous distribution of single mode quadratures $X_k(\varphi) = a_k^\dagger e^{i\varphi} + a_k e^{-i\varphi}$ and their two-mode counterparts $X_\pm(\varphi)$ (X_k and Y_k are used as abbreviations for $X_k(0)$ and $X_k(\pi/2)$). For the squeezed vacuum state, the quadrature distribution of individual modes is isotropic and centred around the origin, reflecting the undefined phase. The correlation of the modes leads to anisotropic and squeezed distributions of the two-mode quadratures $X_\pm(\varphi)$ (refs 10, 20), as seen in Fig. 1a.

Here we report the development of atomic homodyne detection and apply this technique to measure quadratures of twin-atom matter-wave fields generated by controlled spin-changing collisions in a BEC. The experiments involve a spinor BEC of ^{87}Rb trapped in a one-dimensional optical lattice with few hundred atoms per site and a lattice spacing of 5.5 μm (Fig. 1b). A high lattice potential ensures that tunnelling between sites is negligible on the experimental timescale, such that the condensates in the different sites are independent[21]. The density in the centre of the lattice sites is of the order of $2\times10^{14}\,\mathrm{cm}^{-3}$, resulting in a minimal spin healing length of $\xi \approx 1\,\mu\mathrm{m}$, which is comparable to the extension of the on-site wavefunction (approximately 1.1 μm full-width at half-maximum, FWHM). The spatial degrees of freedom are therefore frozen and the dynamics happens exclusively in the hyperfine spin[22]. The condensate is initially prepared in the $(F, m_F) = (2, 0)$ hyperfine state, which serves as the analogue to the pump mode in an OPA. Making use of a combination of quadratic Zeeman shift and state-dependent microwave dressing[23,24], the spin dynamics is energetically restricted to a spin-1 subspace defined by the 0, ±1 Zeeman states (see Supplementary Information). We choose a constant evolution time under spin-changing collisions of 22 ms in all our experiments. In the regime of small pump mode depletion, spin-changing collisions can be described by an OPA-type pair coupling between the Zeeman states[11,12,16,17] with the +1 and −1 states as signal (↑) and idler (↓) modes, respectively. In optics, the integrated nonlinearity is restricted by short interaction times, a limit which is surpassed in the atomic system, such that continuous-variable entangled states with large populations can be realized. As the underlying process generates pairs of atoms in the signal and idler modes—twin atoms—the mean and variance of the population difference $N_- = N_\downarrow - N_\uparrow$ should ideally vanish although the total population $N_+ = N_\downarrow + N_\uparrow$ fluctuates strongly. This can be experimentally confirmed by absorption imaging of the atomic cloud[21] where state selectivity is achieved by Stern-Gerlach separation (Fig. 1b, c).

A quantitative analysis of the variance of the population difference $\Delta^2 N_-$ (where Δ^2 indicates variance) and the distribution of the population sum N_+ gives a first indication of the quantum state of the system (Fig. 2). Owing to the pump mode population dependent

[1]Kirchhoff-Institut für Physik, Universität Heidelberg, Im Neuenheimer Feld 227, 69120 Heidelberg, Germany. [2]Weizmann Institute of Science, Rehovot 76100, Israel. †Present address: Harvard-Smithsonian CfA, Harvard University Department of Physics, Cambridge, Massachusetts 02138, USA.

©2011 Macmillan Publishers Limited. All rights reserved

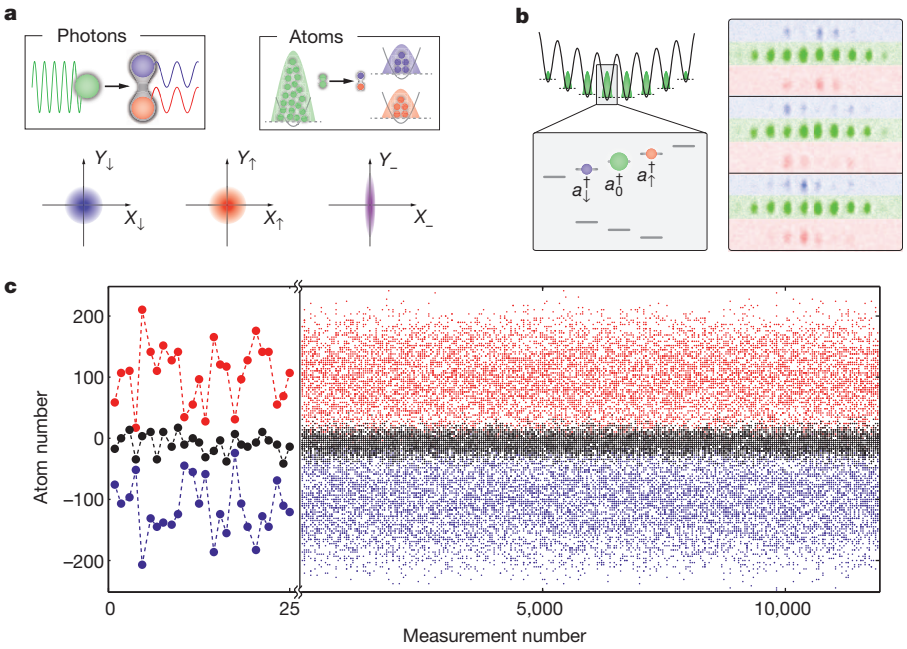


Figure 1 | Analogy to optics and measured population correlations of twin-atom states. **a**, Parametric down-conversion with light, and the analogy to atomic spin-changing collisions. In a nonlinear medium, effective interactions (top) result in the creation of photon pairs in the signal (red) and idler (blue) modes from pump mode photons (green). The quadratures of the output modes (bottom) are centred around the origin—the individual ones are isotropic, while the two-mode quadratures are squeezed reflecting their correlations (purple). See text for nomenclature. Pair creation due to spin-changing collisions in tight traps is an analogous process in quantum atom optics. **b**, Experimental system and exemplary raw absorption pictures. Left, schematic representation (top) of the one-dimensional optical lattice on which is superimposed an optical dipole trap to tightly confine the atoms in all directions. In each site, selected Zeeman levels of the hyperfine spin (bottom) are the only accessible degrees of freedom. The states involved in the spin-changing collision process are highlighted (see text for nomenclature). Right, three typical experimental pictures after spin-changing collisions where the clear spatial separation of sites and states can be seen. **c**, Population correlations. These correlations are visible directly on the raw data (right) where the atom numbers in the signal (red) and idler (blue, numbers shown with opposite sign) modes fluctuate strongly while their difference (black) shows small noise. Left, a zoom into 25 experimental realizations, clearly showing that the fluctuations are common mode.

nonlinearity, the mean atom number in signal and idler mode $\langle N_+\rangle$ increases nonlinearly with the total atom number N, resulting in a growing fraction $\langle N_+\rangle/N$. In the small $N \lesssim 300$ limit, the observed distribution of the total population N_+ matches the prediction for the squeezed vacuum (taking detection noise into account by convolution with a Gaussian) with maximal squeezing parameter $r \approx 2$ (ref. 10). The corresponding mean population is $\langle N_+\rangle \approx 25$ and the data shows a comparably large standard deviation—in contrast, the noise in the population difference N_- almost vanishes ($\Delta N_- < 1.9$ atoms). It is important to note that the atomic variances reported throughout this Letter are corrected for detection noise, as detailed in the Supplementary Information. For larger total atom numbers N,

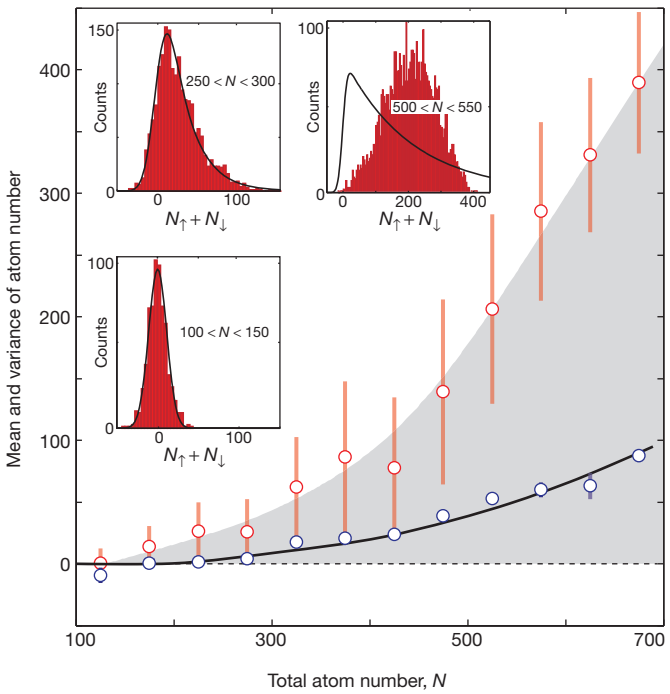


Figure 2 | Population of the signal and idler modes. Main figure: red data, mean population $\langle N_+\rangle = \langle N_\uparrow + N_\downarrow\rangle$ versus total atom number N. Error bars, fluctuations (1 s.d.) of the data. The large uncertainty is a feature of the nonlinear process itself, which leads to a non-Gaussian distribution of the mode population. Grey area, the sub-Poisson regime, the upper bound of which is obtained by a spline fit to $\langle N_+\rangle$. Blue data points, variance of the population difference $\Delta^2 N_-$, which is compatible with zero for small N (error bars, 1 s.d. statistical uncertainty). Its increase with N is reproduced when taking particle loss due to spin relaxation into account (black line). Insets, the distribution of the mode population for indicated total atom numbers. For small N, the distributions match the prediction for a non-depleted pump with the measured mean population $\langle N_+\rangle$ (black lines) whereas for larger N they clearly differ. The fitted squeezed vacuum distribution for $250 < N < 300$ corresponds to a squeezing parameter of $r \approx 2$ (ref. 10).

©2011 Macmillan Publishers Limited. All rights reserved

the pump depletion can no longer be neglected[22]; this results in a breakdown of the analogy to the OPA picture, and the observed distributions of N_+ differ from the squeezed vacuum (Fig. 2 insets). In this regime, the well controllable spin dynamics offer prospects for the deterministic generation of non-gaussian entangled twin-atom states[25]. Density dependent two-body spin relaxation loss causes the perfect pair correlations to deteriorate, such that the relative population noise grows with N. Averaging the results for all total atom numbers, we find the noise $\Delta^2 N_-$ suppressed by 6.9 dB below the Poisson level $\langle N_+ \rangle$.

The observable populations do not suffice to characterize the two-mode state; we must access coherences between the modes, the two-mode quadratures. We use atomic homodyning by analogy to the experimental technique used in optics, whereby the quadratures are measured relative to a reference field. Yet in the atomic case, the population of available local oscillator (LO) reference fields is often limited to rather small atom numbers. This has important consequences for the observable quadratures $\tilde{X}_k = \left(a_k^\dagger a_{\mathrm{LO}} + a_k a_{\mathrm{LO}}^\dagger\right) \Big/ \sqrt{\left\langle a_{\mathrm{LO}}^\dagger a_{\mathrm{LO}} \right\rangle}$, which in general differ from the canonical ones. A measure for this difference is given by the ratio of the quantum mode population to the local oscillator population, by which the Heisenberg uncertainty relation is modified[26], $\Delta\tilde{X}_k \Delta\tilde{Y}_k \geq 1 - \langle N_k \rangle \Big/ \left\langle a_{\mathrm{LO}}^\dagger a_{\mathrm{LO}} \right\rangle$.

For the twin-atom state, we expect peculiar coherence properties: the quadratures of the individual modes fluctuate strongly around a vanishing mean, while the two-mode quadratures $X_+(\varphi)$ show reduced fluctuations for certain φ reflecting the quantum correlations. To reveal these characteristics in the coherences of the generated quantum fields, we employ a homodyning scheme such that information about the two-mode quadratures $\tilde{X}_+(\varphi)$ can be obtained. Measurement of these observables requires comparison of each mode with a local oscillator. We ensure relative local oscillator phase stability by using the pump mode (2, 0) as the simultaneous reference, whose phase φ can be controlled by microwave dressing. An atomic three-port beam splitter is realized by radio-frequency coupling of signal and idler modes to the pump. Analysis of the fluctuations in atom number sum and difference $N_{(2,-1)} \pm N_{(2,1)}$ between the (2, −1) and (2, 1) Zeeman states after the mixing provides, after proper normalization, an upper bound (ub) for the variance of the two-mode quadratures $\Delta^2 \tilde{X}_\pm^{\mathrm{ub}}(\varphi)$. The measured fluctuations include further noise contributions governed by population fluctuations in the signal and idler modes, which are small only for the quadrature difference. Further details of the analysis in the three mode framework and the choice of the normalizations can be found in Supplementary Information. Figure 3a and b shows raw data in the regime of small pump depletion of the measured normalized atom number difference and sum versus local oscillator phase φ. No phase dependence is visible in the mean, but nevertheless, the fluctuations are strongly modulated, indicating phase correlations between the two atomic modes.

We set up a different experiment to obtain the quadrature fluctuations of the signal and idler modes individually. As local oscillators, we prepare an atomic population in the $(1, \mp 1)$ states before initiating the dynamics. Mixing of the local oscillators with the quantum modes is done by a two-photon microwave and radio-frequency $\pi/2$ coupling pulse. For both the signal and idler mode, we find coupling phase independent fluctuations of the atomic quadratures $\tilde{X}_{\uparrow(\downarrow)} = \left(N_{(2,\pm 1)} - N_{(1,\mp 1)}\right) \Big/ \sqrt{\langle N_{\mathrm{LO}} \rangle}$ and a vanishing mean. The mean local oscillator population $\langle N_{\mathrm{LO}} \rangle$ of the $(1, \mp 1)$ state was measured by omitting the mixing pulse.

To ensure that our observations are indeed resulting directly from the spin-changing collision process, the technical phase noise has to be characterized. The one-dimensional optical lattice set-up allows for this characterization by a reference measurement in which the population in the $(2, \pm 1)$ states is also prepared by electromagnetic coupling and spin-changing collisions are tuned off-resonance during the evolution time. The inter-site correlations observed after the mixing are shown in Fig. 3c for this reference experiment (top) and for the quantum state prepared by spin changing collisions (bottom). In both cases, we observe similarly large on-site fluctuations in the atom

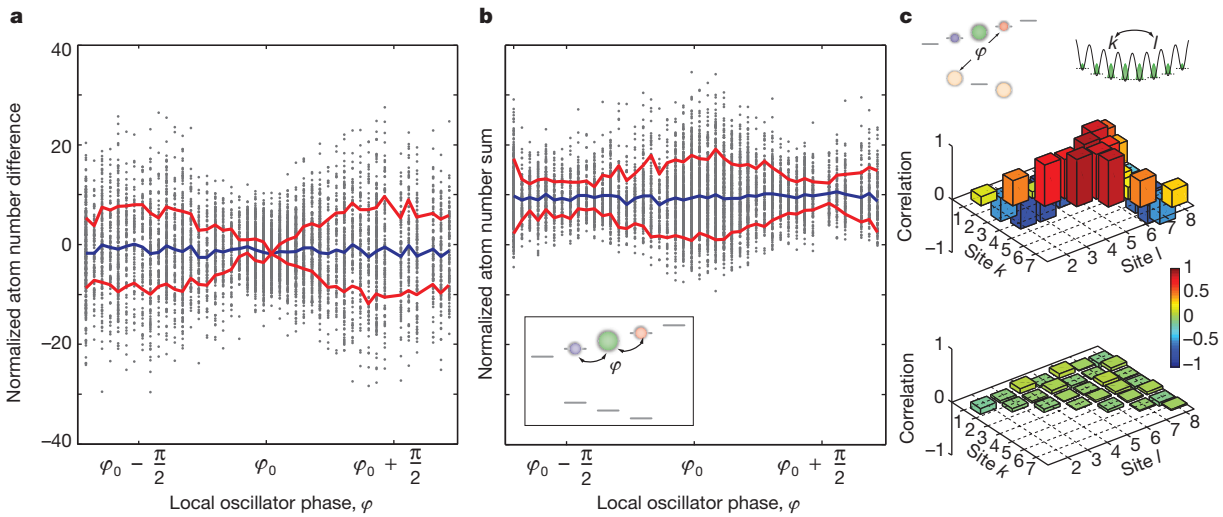


Figure 3 | Atomic homodyning. **a**, Homodyne measurement of the quadrature difference. After symmetric coupling to the $m_F = 0$ state (inset in **b**), the normalized population difference in the $m_F = \pm 1$ states provides a measure for the two-mode quadrature difference. Phase dependent noise is visible directly on the raw data while no dependence is observed in the mean population difference (solid blue line). **b**, Homodyne measurement of the quadrature sum. An upper bound for the variance of the two-mode quadrature sum is obtained from the fluctuations of the normalized population sum after the coupling. The fluctuations of the raw data are strongly phase dependent while the mean (blue solid line) stays constant. The variance minimum is shifted by $\pi/2$ as compared to **a**, indicating noise suppression in the orthogonal two-mode quadrature. In **a** and **b**, the data correspond to all measurements with total numbers $150 < N < 200$, and the red lines indicate the inferred standard deviation above and below the mean after subtraction of the readout noise. **c**, Classical versus quantum preparation—inter-site correlation measurement. Characterization of classical phase noise is crucial to obtain information about the individual quadratures of the signal and idler modes. For initial states prepared classically, that is by linear coupling, strong inter-site correlations between different sites k,l in the optical lattice after mixing with the local oscillator (top illustrations) reveal classically washed out coherences (middle). In contrast, when the $(F = 2, m_F = \pm 1)$ states are populated by spin-changing collisions, no correlations are observed within the statistical uncertainty showing that $\langle \tilde{X}_{\uparrow(\downarrow)} \rangle = 0$ as a direct result of the spin dynamics (bottom). The colour scale is used to represent the observed inter-site correlations.

©2011 Macmillan Publishers Limited. All rights reserved

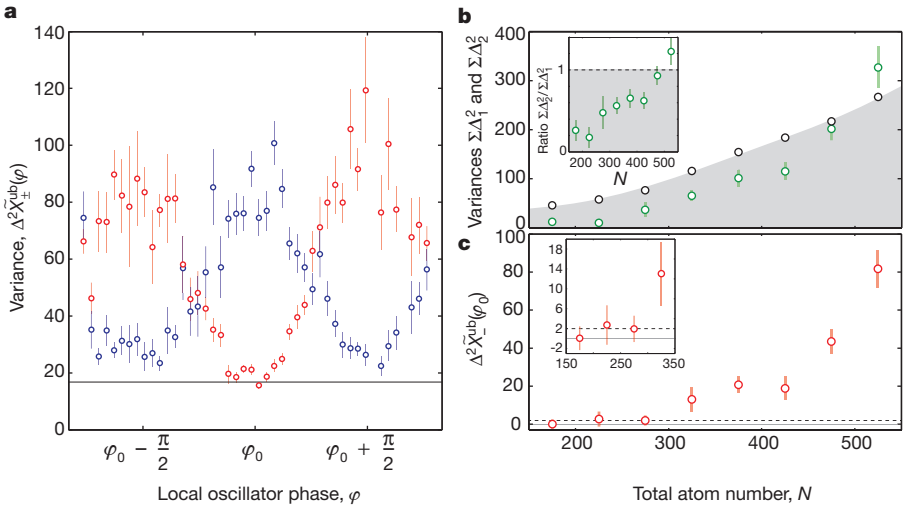


Figure 4 | Two-mode quadrature fluctuations and mode inseparability. **a**, Two-mode sum and difference quadratures. The measured variances $\Delta^2\tilde{X}^{ub}_{+}(\varphi)$ (blue) and $\Delta^2\tilde{X}^{ub}_{-}(\varphi)$ (red)—upper bounds for the sum and difference two-mode quadrature variances—are plotted versus local oscillator phase for total atom numbers $150 < N < 200$. The shown experimental results are not corrected for readout noise (in contrast to the values stated throughout the text and in **b**, **c**) but its magnitude is shown as the black line. **b**, Continuous-variable quadrature entanglement. Inseparability is detected in the grey shaded region where $\sum\Delta_2^2$, the upper bound for the sum of the two-mode quadrature variance minima (green), is smaller than $\sum\Delta_1^2$, the sum of the variance in the individual quadratures (black). Inset, ratio $\sum\Delta_2^2/\sum\Delta_1^2$. **c**, Minimum of the two-mode difference quadrature fluctuations. For small total atom numbers N—in the non-depleted pump regime and where spin relaxation loss is small—the minimum two-mode quadrature variance $\Delta^2\tilde{X}^{ub}_{-}(\varphi_0)$ inferred from a local quadratic fit is at the noise limit for two coherent fields (dashed black line). Inset, zoom into the small N region. All error bars, 1-s.d. uncertainties.

number difference. However, in the reference experiment we can use the inter-site correlations to reduce these fluctuations by a factor of 20, such that the remaining fluctuations are close to the expected noise limit for two coherent modes. This shows that inter-site correlations detect finite coherences that have been washed out by classical phase noise in our measurements. In the case of spin-changing collisions, no correlations are observed within the statistical uncertainty, meaning that the vanishing coherences are indeed caused by the process itself. Given this result, we can be sure that technical phase noise does not influence the sum of the orthogonal quadrature variances $\Delta^2\tilde{X}_{\uparrow(\downarrow)} + \Delta^2\tilde{Y}_{\uparrow(\downarrow)} = 2\Delta^2\tilde{X}_{\uparrow(\downarrow)}$ (the measured quadrature variance is independent of the local oscillator phase), such that it is a useful observable to characterize the quadrature fluctuations. For further details on the inter-site correlation measurement, see Supplementary Information.

Inter-mode entanglement of the twin-atom state is reflected in suppressed noise in the two-mode quadratures compared to the quadrature fluctuations of the individual modes[13] $\Delta^2\tilde{X}_- + \Delta^2\tilde{Y}_+ < \Delta^2\tilde{X}_\uparrow + \Delta^2\tilde{X}_\downarrow + \Delta^2\tilde{Y}_\uparrow + \Delta^2\tilde{Y}_\downarrow$. When investigating this inequality in our experiment, care has to be taken to ensure comparability of the single- and two-mode measurements[26]. In both measurements, the ratio of the sum of the population in the signal and idler modes to the reference mode population is below 10% for small $N < 200$, and grows to approximately 55% for the largest N. In Fig. 4a we plot the measured upper bound for the two-mode quadrature variances versus local oscillator phase $\Delta^2\tilde{X}^{ub}_{\pm}(\varphi)$ for an exemplary total atom number. The sum of the minima at φ_0 and $\varphi_0 - (\pi/2)$ provides an upper bound for the sum of the orthogonal two-mode quadrature variances $\sum\Delta_2^2 = \Delta^2\tilde{X}^{ub}_{-} + \Delta^2\tilde{Y}^{ub}_{+}$, while the measurement of the single-mode quadrature variances $\sum\Delta_1^2 = 2\left(\Delta^2\tilde{X}_\uparrow + \Delta^2\tilde{X}_\downarrow\right)$ has been discussed above. The generated quantum states fulfil the inequality $\sum\Delta_2^2 < \sum\Delta_1^2$ for a wide range of total atom numbers (Fig. 4b), showing that the produced twin-atom state is inseparable—the observed minimal ratio is $\sum\Delta_2^2/\sum\Delta_1^2 \approx 0.2$.

For the squeezed vacuum, which is theoretically expected in the non-depleted pump regime, the two-mode quadrature variances should be reduced below the noise level of two coherent modes $\Delta^2\tilde{X}^{ub}_{-}(\varphi) = 2$. For the more precisely detectable variance of the quadrature difference, we find the noise minimum for small total N to be comparable to this level, limited by the experimental signal to noise ratio. In Fig. 4c we show the dependence of this noise minimum on the total atom number N—detection noise subtraction is crucial here, and without it we find $\Delta^2\tilde{X}^{ub}_{-}(\varphi) \approx 17$ as the minimal value.

In conclusion, we have developed an atomic homodyne detection method, which allows for the measurement of quadrature correlations of twin-atom quantum fields. The continuous-variable entangled states have been produced in a deterministic manner using controlled atomic spin interactions—for small total atom numbers, the observed population distributions are compatible with the atomic two-mode squeezed vacuum. Einstein–Podolsky–Rosen (EPR) entanglement is revealed by $\Delta^2(X_\uparrow - X_\downarrow)\Delta^2(Y_\uparrow + Y_\downarrow) < 1$, and we find 4 ± 17 for this value, showing that EPR entanglement for atomic quadratures is within reach. In a future extension of our scheme, the local oscillator might be split before the mixing, enabling precise detection of the two-mode quadrature sum as well[26]. Spatial separation of the modes can be implemented by employing the Stern-Gerlach technique, as an alternative approach to recently reported experiments[27,28]. The regime of large average pair population is accessible, such that highly non-classical quantum states might be generated by controlled quantum spin evolution[25].

We note that in parallel to this work, sub-Poissonian number fluctuations after spin-changing collisions have been observed by two other groups[29,30], and it has been shown that the generated quantum correlated states are useful for noise interferometry beyond the standard quantum limit[30].

Received 4 March; accepted 18 October 2011.
Published online 30 November 2011.

1. Einstein, A., Podolsky, B. & Rosen, N. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47,** 777–780 (1935).
2. Giovannetti, V., Lloyd, S. & Maccone, L. Quantum-enhanced measurements: beating the standard quantum limit. *Science* **306,** 1330–1336 (2004).
3. Braunstein, S. L. & van Loock, P. Quantum information with continuous variables. *Rev. Mod. Phys.* **77,** 513–577 (2005).
4. Braunstein, S. L. & Kimble, H. J. Teleportation of continuous quantum variables. *Phys. Rev. Lett.* **80,** 869–872 (1998).
5. Opatrný, T. & Kurizki, G. Matter-wave entanglement and teleportation by molecular dissociation and collisions. *Phys. Rev. Lett.* **86,** 3180–3183 (2001).

©2011 Macmillan Publishers Limited. All rights reserved

6. Hammerer, K., Sørensen, A. S. & Polzik, E. S. Quantum interface between light and atomic ensembles. *Rev. Mod. Phys.* **82,** 1041–1093 (2010).
7. Reid, M. D. *et al.* The Einstein-Podolsky-Rosen paradox: from concepts to applications. *Rev. Mod. Phys.* **81,** 1727–1751 (2009).
8. Ou, Z. Y., Pereira, S. F., Kimble, H. J. & Peng, K. C. Realization of the Einstein-Podolsky-Rosen paradox for continuous variables. *Phys. Rev. Lett.* **68,** 3663–3666 (1992).
9. Reid, M. D. Demonstration of the Einstein-Podolsky-Rosen paradox using nondegenerate parametric amplification. *Phys. Rev. A* **40,** 913–923 (1989).
10. Walls, D. & Milburn, G. *Quantum Optics* (Springer, 2008).
11. Duan, L.-M., Sørensen, A., Cirac, J. I. & Zoller, P. Squeezing and entanglement of atomic beams. *Phys. Rev. Lett.* **85,** 3991–3994 (2000).
12. Pu, H. & Meystre, P. Creating macroscopic atomic Einstein-Podolsky-Rosen states from Bose-Einstein condensates. *Phys. Rev. Lett.* **85,** 3987–3990 (2000).
13. Raymer, M. G., Funk, A. C., Sanders, B. C. & de Guise, H. Separability criterion for separate quantum systems. *Phys. Rev. A* **67,** 052104 (2003).
14. Julsgaard, B., Kozhekin, A. & Polzik, E. S. Experimental long-lived entanglement of two macroscopic objects. *Nature* **413,** 400–403 (2001).
15. Bloch, I., Dalibard, J. & Zwerger, W. Many-body physics with ultracold gases. *Rev. Mod. Phys.* **80,** 885–964 (2008).
16. Leslie, S. R. *et al.* Amplification of fluctuations in a spinor Bose-Einstein condensate. *Phys. Rev. A* **79,** 043631 (2009).
17. Klempt, C. *et al.* Parametric amplification of vacuum fluctuations in a spinor condensate. *Phys. Rev. Lett.* **104,** 195303 (2010).
18. Chang, M.-S. *et al.* Observation of spinor dynamics in optically trapped ^{87}Rb Bose-Einstein condensates. *Phys. Rev. Lett.* **92,** 140403 (2004).
19. Schmaljohann, H. *et al.* Dynamics of $F = 2$ spinor Bose-Einstein condensates. *Phys. Rev. Lett.* **92,** 040402 (2004).
20. Caves, C. M. & Schumaker, B. L. New formalism for two-photon quantum optics. I. Quadrature phases and squeezed states. *Phys. Rev. A* **31,** 3068–3092 (1985).
21. Gross, C., Zibold, T., Nicklas, E., Estève, J. & Oberthaler, M. K. Nonlinear atom interferometer surpasses classical precision limit. *Nature* **464,** 1165–1169 (2010).
22. Law, C., Pu, H. & Bigelow, N. Quantum spins mixing in spinor Bose-Einstein condensates. *Phys. Rev. Lett.* **81,** 5257–5261 (1998).
23. Kronjäger, J., Becker, C., Navez, P., Bongs, K. & Sengstock, K. Magnetically tuned spin dynamics resonance. *Phys. Rev. Lett.* **97,** 110404 (2006).
24. Gerbier, F., Widera, A., Fölling, S., Mandel, O. & Bloch, I. Resonant control of spin dynamics in ultracold quantum gases by microwave dressing. *Phys. Rev. A* **73,** 041602 (2006).
25. Diener, R. B. & Ho, T. Quantum spin dynamics of spin-1 Bose gas. Preprint at ⟨http://arXiv.org/abs/cond-mat/0608732v1⟩ (2006).
26. Ferris, A. J., Olsen, M. K., Cavalcanti, E. G. & Davis, M. J. Detection of continuous variable entanglement without coherent local oscillators. *Phys. Rev. A* **78,** 060104 (2008).
27. Jaskula, J.-C. *et al.* Sub-Poissonian number differences in four-wave mixing of matter waves. *Phys. Rev. Lett.* **105,** 190402 (2010).
28. Bücker, R. *et al.* Twin-atom beams. *Nature Phys.* **7,** 608–611 (2011).
29. Bookjans, E. M., Hamley, C. D. & Chapman, M. S. Strong quantum spin correlations observed in atomic spin mixing. Preprint at ⟨http://arXiv.org/abs/1109.2185⟩ (2011).
30. Lücke, B. *et al.* Twin matter waves for interferometry beyond the classical limit. *Science* doi:10.1126/science.1208798 (published online 13 October 2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge discussions with P. Grangier, A. Aspect, A. J. Ferris, M. J. Davis and B. C. Sanders. This work was supported by the Forschergruppe FOR760, the Deutsche Forschungsgemeinschaft, the German–Israeli Foundation, the Heidelberg Center for Quantum Dynamics, Landesstiftung Baden-Württemberg, the ExtreMe Matter Institute and the European Commission Future and Emerging Technologies Open Scheme project MIDAS (Macroscopic Interference Devices for Atomic and Solid-State Systems). G.K. acknowledges support from the Humboldt-Meitner Award and the Deutsche-Israelische Projektgruppe (DIP).

Author Contributions N.B.-G. and G.K. contributed to the formulation of the problem. C.G., H.S., E.N., T.Z. and M.K.O. contributed equally to the study. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.K.O. (homodyning@matterwave.de).

©2011 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature10647

Trifluoromethylation of arenes and heteroarenes by means of photoredox catalysis

David A. Nagib[1] & David W. C. MacMillan[1]

Modern drug discovery relies on the continual development of synthetic methodology to address the many challenges associated with the design of new pharmaceutical agents[1]. One such challenge arises from the enzymatic metabolism of drugs *in vivo* by cytochrome P450 oxidases, which use single-electron oxidative mechanisms to rapidly modify small molecules to facilitate their excretion[2]. A commonly used synthetic strategy to protect against *in vivo* metabolism involves the incorporation of electron-withdrawing functionality, such as the trifluoromethyl (CF_3) group, into drug candidates[3]. The CF_3 group enjoys a privileged role in the realm of medicinal chemistry because its incorporation into small molecules often enhances efficacy by promoting electrostatic interactions with targets, improving cellular membrane permeability, and increasing robustness towards oxidative metabolism of the drug[4–6]. Although common pharmacophores often bear CF_3 motifs in an aromatic system, access to such analogues typically requires the incorporation of the CF_3 group, or a surrogate moiety, at the start of a multi-step synthetic sequence. Here we report a mild, operationally simple strategy for the direct trifluoromethylation of unactivated arenes and heteroarenes through a radical-mediated mechanism using commercial photocatalysts and a household light bulb. We demonstrate the broad utility of this transformation through addition of CF_3 to a number of heteroaromatic and aromatic systems. The benefit to medicinal chemistry and applicability to late-stage drug development is also shown through examples of the direct trifluoromethylation of widely prescribed pharmaceutical agents.

The CF_3 moiety is typically installed by means of cross-coupling technologies catalysed by transition metals. These methods historically required the stoichiometric use of metal salts or organometallic complexes, and have shown limited generality[7]. Recently, however, copper and palladium complexes have been successfully shown to enable this transformation in a catalytic fashion[8]. Using either a nucleophilic or electrophilic source of CF_3 (Ruppert–Prakash or Umemoto's reagent, respectively), it is now possible to install an aryl trifluoromethyl moiety in place of halides and boronic acids, or adjacent to tailored directing groups (Fig. 1b)[9–12]. These cross-coupling reactions have facilitated the synthesis of a range of CF_3 analogues without the reliance on prefluorinated building blocks. However, the ideal position of CF_3 substitution on a drug candidate must still be determined through parallel, multi-step syntheses employing aryl precursors bearing activating groups at various positions around an aromatic ring. Therefore, a remaining challenge in the synthesis of organofluorine drugs is this reliance on substituted precursors and their varying amenability to cross-coupling reactions.

To address this challenge, we reasoned that the direct installation of a CF_3 group to a drug candidate—ideally at the position(s) of metabolic susceptibility—would obviate the need for pre-functionalization. Thus, this direct trifluoromethylation of aryl and heteroaryl compounds would, in a single chemical operation, preclude the need for redundant synthetic efforts and simultaneously protect against metabolic oxidation *in vivo*. In the first part of our design plan, we proposed that a single-electron pathway might provide such an approach to address the unsolved challenge of trifluoromethylation of non-activated arenes (Fig. 1c). Specifically, we sought to take advantage of photoredox catalysis, which provides a mild and efficient method for accessing electrophilic radicals, such as $^{\bullet}CF_3$, by means of photosynthesis-inspired redox chemistry[13–15]. This mode of reactivity employs polypyridyl organometallic complexes, such as $Ru(phen)_3Cl_2$ (**1** in Fig. 2, phen = phenanthroline), whose excitation by visible light (for example, a household light bulb) at room temperature provides a strongly oxidizing or reducing catalyst that can rapidly engage a variety of substrates to generate high energy, reactive species[16].

As a second part of our design plan, we recognized that the moderate to low nucleophilicity of many classes of unfunctionalized arenes has been typically underemployed in substitution reactions, presumably owing to the harsh conditions required to manipulate such aromatic rings via two-electron pathways (for example, nucleophilic or electrophilic aromatic substitution). As an alternative, we considered that a strategy based on radical addition could exploit this large area of aromatic structural space, thereby rendering a more general approach to CF_3 incorporation. Our strategy, outlined in Fig. 1, involves site-specific incorporation of electrophilic radicals at metabolically susceptible positions of arenes and heteroarenes—in analogy to the single-electron aryl modification processes employed by enzymes. This approach precludes the need for pre-functionalization of arenes, and offers a complementary method of accessing aryl trifluoromethyl functionality from late-stage synthetic intermediates.

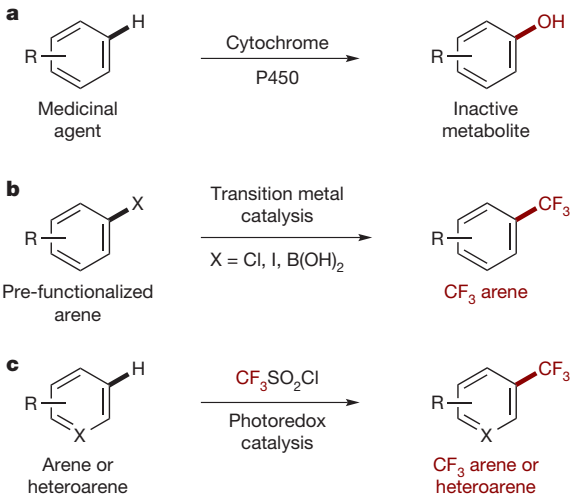


Figure 1 | Direct trifluoromethylation of aryl and heteroaryl C–H bonds. The excretion of medicinal agents is facilitated by remote functionalization of aromatic moieties (**a**). A common medicinal chemistry approach to block this catabolism involves cross-coupling of CF_3 reagents to pre-functionalized arenes (**b**). In analogy to enzymatic processes, our photoredox strategy (**c**) enables the direct C–H functionalization of unfunctionalized arenes and heteroarenes with the CF_3 pharmacophore using a cheap and easy to handle $^{\bullet}CF_3$ source.

[1]Merck Center for Catalysis at Princeton University, Department of Chemistry, 192 Frick Laboratory, Princeton, New Jersey 08540, USA.

©2011 Macmillan Publishers Limited. All rights reserved

Building on our recent investigations of the α-trifluoromethylation of carbonyls with CF_3I as a $^{\bullet}CF_3$ source[17], our initial exploration of a photoredox strategy towards the development of this transformation was successful within the limited context of π-rich arenes. We quickly observed, however, that reaction efficiency was diminished by competitive aryl iodination, presumably as a result of CF_3–I homolysis. Our mechanistic understanding of this undesired pathway—wherein initiation of the photocatalytic cycle relies on sacrificial oxidation of an electron-rich arene—prompted us to seek an alternative method of catalytic initiation. Because direct reduction of CF_3I (at a potential of −1.52 V versus the saturated calomel electrode, SCE, in cyclic voltammetry)[18] by visible-light-excited photocatalyst $^*Ru(phen)_3^{2+}$ **2** (Fig. 2; −0.90 V versus SCE)[19] is thermodynamically challenging, we questioned whether the incorporation of an electron-deficient system such as SO_2 within CF_3X might enable more facile reduction[20]. Our investigations led us to the use of trifluoromethanesulphonyl chloride (known as triflyl chloride, TfCl), which has been employed in a non-general sense to effect aryl trifluoromethylation of benzene and other simple arenes via atom transfer catalysis[21–23]. In particular, the relative cost and ease of handling of TfCl in comparison to other $^{\bullet}CF_3$ sources rendered this a highly attractive reagent.

For the proposed photoredox catalytic cycle, we assumed that single-electron transfer (SET) reduction of triflyl chloride should be concurrent with oxidation of $^*Ru(phen)_3^{2+}$ **2** to $Ru(phen)_3^{3+}$ **3** (Fig. 2). The ensuing CF_3SO_2Cl radical anion should then rapidly collapse to generate the stabilized $^{\bullet}CF_3$, a process that should be entropically driven by the release of SO_2 and chloride[24]. This electron deficient trifluoromethyl radical is then well-suited to add—in a selective fashion—to the most electron-rich position of any arene or heteroarene[25,26]. In the case of benzene, the resultant cyclohexadienyl radical **4** (Fig. 2; −0.1 V versus SCE)[27] should then undergo a second SET event with the now strongly oxidizing Ru^{3+} photocatalyst **3** (Fig. 2; +1.31 V versus SCE) to regenerate ground-state photocatalyst **1**.

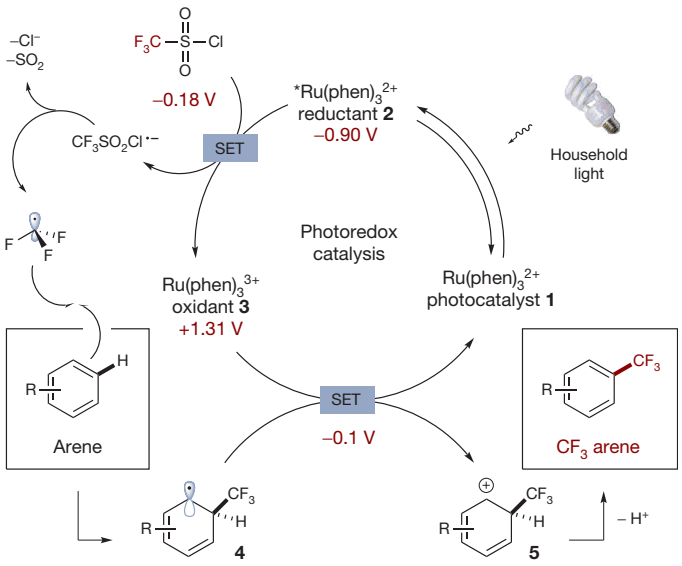


Figure 2 | Proposed mechanism for the direct trifluoromethylation of aryl C–H bonds via photoredox catalysis. The photoredox catalytic cycle is initiated via excitation of photocatalyst **1** to excited state **2** with a household light bulb. Subsequent reduction of triflyl chloride by reductant **2** (via single electron transfer, that is, SET) provides oxidant **3** along with the radical anion of triflyl chloride. This high energy species spontaneously collapses to form CF_3 radical (top left), which selectively combines with aromatic systems enabling direct CF_3 substitution. Catalyst **3**-promoted oxidation of the ensuing radical **4** completes the catalytic cycle and provides cyclohexadienyl cation **5**, whose facile deprotonation provides the desired CF_3 arene. This mechanism is consistent with measured redox values (shown below respective compounds, in maroon).

Finally, facile deprotonation of the cyclohexadienyl cation **5** (Fig. 2) with a suitable base would provide the desired trifluoromethyl arene in a redox, catalytic fashion without the need for arene pre-functionalization.

As illustrated in Fig. 3, we found that this new photoredox protocol allows the direct incorporation of CF_3 into a broad range of arene and heteroarene rings using only TfCl and a household light bulb. In contrast to atom transfer pathways for trifluoromethylation that require high temperature (120 °C), our photoredox strategy is successful at room temperature, allowing efficient production of the trifluoromethyl radical. Our initial studies revealed that exposure of five-atom heterocycles to a household 26-W fluorescent light source in the presence of triflyl chloride, $Ru(phen)_3Cl_2$ (**1**) photocatalyst, and base, enabled trifluoromethylation with excellent levels of efficiency (Fig. 3a). More specifically, the fluoroalkylation of pyrroles, furans and thiophenes occurred with excellent selectivity and yield (compounds **6**–**12**, 78–94% yield). Our rationale for the observed selectivity at the C2 position of these heteroaromatics relies on the formation of the conjugated (rather than cross-conjugated) radical and cationic intermediates. We also found that both mono- and bis-trifluoromethylation could be obtained by varying the amount of CF_3 source used (**8** and **9**, 88 and 91% yield, respectively). Additionally, a wide range of ring substitution and heteroatom protecting groups are well tolerated (**10**–**17**, 70–87% yield). Electron-deficient five-atom heterocycles such as thiazoles are also amenable to this trifluoromethylation protocol, and useful levels of efficiency are obtained by increasing equivalents of the CF_3 source (**14**, 70% yield).

We next turned our attention to six-atom heterocycles with the knowledge that there is an unmet need for CF_3 cross-coupling technologies for these valuable pharmacophores. As shown in Fig. 3b, we found that an array of pyrazine, pyrimidine, pyridine and pyrone substrates are compatible with this new trifluoromethylation approach (**18**–**32**, 70–94% yield). We further observed that radical addition and the ensuing oxidation of electronically deficient halide-substituted heteroaromatics (such as 2,6-dichloropyrazine) were also operative. Unsubstituted analogues of each of these heterocycles can also be trifluoromethylated, albeit with somewhat diminished efficiency. Notably, the commercial photocatalyst $Ir(Fppy)_3$ (Fppy = 2-(2,4-difluorophenyl)pyridine, see Supplementary Information for commercial sources, synthesis and photophysical properties) is capable of providing higher levels of reactivity (presumably due to a longer-lived excited state), although lower levels of regiocontrol are observed with this system (the role of catalyst reduction potentials in these outcomes is being explored). Finally, the trifluoromethylation of arenes (shown in Fig. 3c) demonstrates the complementary reactivity offered by this radical process. *Ortho*-trifluoromethyl containing anilines, anisoles, thioanisoles and xylenes—typically challenging motifs to access via cross-coupling strategies—are also readily prepared using this methodology (**34**–**40**, 70–84% yield). Tolerance of halides, multiple substitution and large steric bulk are additional benefits of this radical-based process (**38**–**47**, 72–92% yield).

Our mechanistic hypothesis for the photoredox process described above is supported by emission quenching experiments as well as cyclic voltammetry. In the first case, no quenching interactions are observed between excited state **2** (Fig. 2) and a range of electronically diverse arenes and heteroarenes that are competent partners in this transformation; whereas triflyl chloride exhibits significant fluorescence quenching (Stern–Volmer constant, 22 M^{-1}; Supplementary Fig. 1). Furthermore, we have determined the oxidation potential of triflyl chloride (–0.18 V versus SCE; Supplementary Fig. 2) to be in accord with the proposed reduction step using excited state **2**, as shown in Fig. 2. It is also noteworthy that decreased reactivity is observed in the absence of base, presumably due to HCl formation in the reaction. Inorganic bases are preferred to organic bases, as pyridines are trifluoromethylated and alkyl amines are prone to oxidative decomposition.

Ultimately, a major benefit of our mild, visible-light-induced trifluoromethylation procedure is its amenability to late-stage synthetic

©2011 Macmillan Publishers Limited. All rights reserved

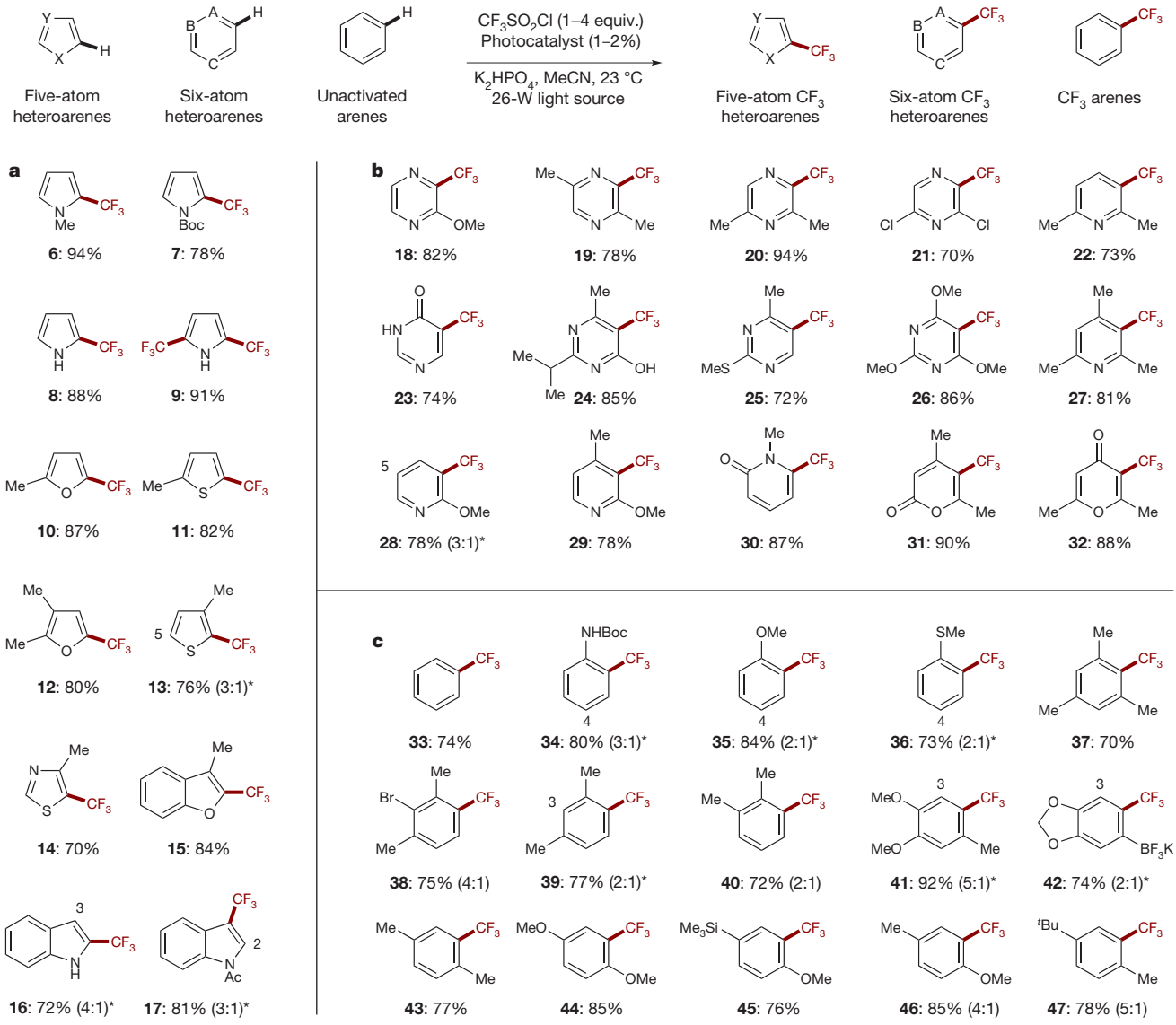


Figure 3 | Radical trifluoromethylation of five- and six-membered heteroarenes and C–H arenes via photoredox catalysis. Aromatic and heteroaromatic systems with varying stereoelectronics are efficiently trifluoromethylated under these standard conditions (top, generalized reaction). Substrate scope includes electron-rich five-atom heteroarenes (**a**), electron-deficient six-atom heteroarenes (**b**), and unactivated arenes (**c**). Isolated yields are indicated below each entry (^{19}F NMR yields for volatile compounds). See Supplementary Information for experimental details. Abbreviations: X, Y, A, B, C represent either C, N, O, or S; MeCN, acetonitrile; Me, methyl; Boc, *tert*-butoxycarbonyl; Ac, acyl; *t*Bu, *tert*-butyl. *The minor regioisomeric position is labelled with the respective carbon atom number.

applications. To demonstrate this potential, biologically active molecules (Fig. 4) were subjected to our photoredox trifluoromethylation protocol to obtain their CF_3 analogues. In some cases, highly regioselective addition is observed, presumably corresponding to positions of high electron density and potential metabolic susceptibility (Fig. 4a). In medicinal chemistry, the simultaneous identification and blocking of such positions with a CF_3 group may prove quite useful. For instance, a DNA base (uracil) analogue, an Alzheimer's drug (Aricept) precursor, and a vitamin (flavone) were each selectively trifluoromethylated at a single position with excellent efficiency (85–94% yield). The first example represents a facile, one-step synthesis of a known CF_3-uracil derivative from its abundant non-fluoro analogue. These uridines have been extensively studied owing to their anti-viral and cancer-treating properties[28].

Alternatively, a successful drug discovery programme is likely to eliminate potential sites of metabolic reactivity from a medicinal agent during the course of its development[29]. Without an especially electron-rich π-system present, competitive trifluoromethylation of the remaining aromatic sites should be observed. Not surprisingly, radical trifluoromethylation of lidocaine and ibuprofen results in alkylation at both of their available aromatic positions (Fig. 4b). Subjection of the cholesterol-lowering statin, Lipitor, to our protocol similarly results in promiscuous functionalization at three positions (Fig. 4c; 74% yield, 1:1:1). In the context of drug discovery, the isolation of these separate isomers offers the potential for rapid testing of fluorinated analogues for biological activity from a single late-stage synthetic product. This streamlined process represents an attractive alternative to multi-step, parallel syntheses. Perhaps most importantly, this technology provides a strategically new approach to the introduction of metabolism-blocking fluoroalkyl groups at the end of a medicinal chemistry development effort.

In this work, we have introduced a photoredox-based method that allows for facile trifluoromethylation of aromatic and heteroaromatic systems without the need for aryl ring pre-activation. The value of this transformation has been highlighted via the trifluoromethylation of biologically active molecules with commercially available photocatalysts and a household light bulb. While this manuscript was under review, a complementary method for heterocyclic trifluoromethylation was reported by others[30]. Their approach also employs radical intermediates to harness innate chemical reactivity of unactivated

©2011 Macmillan Publishers Limited. All rights reserved

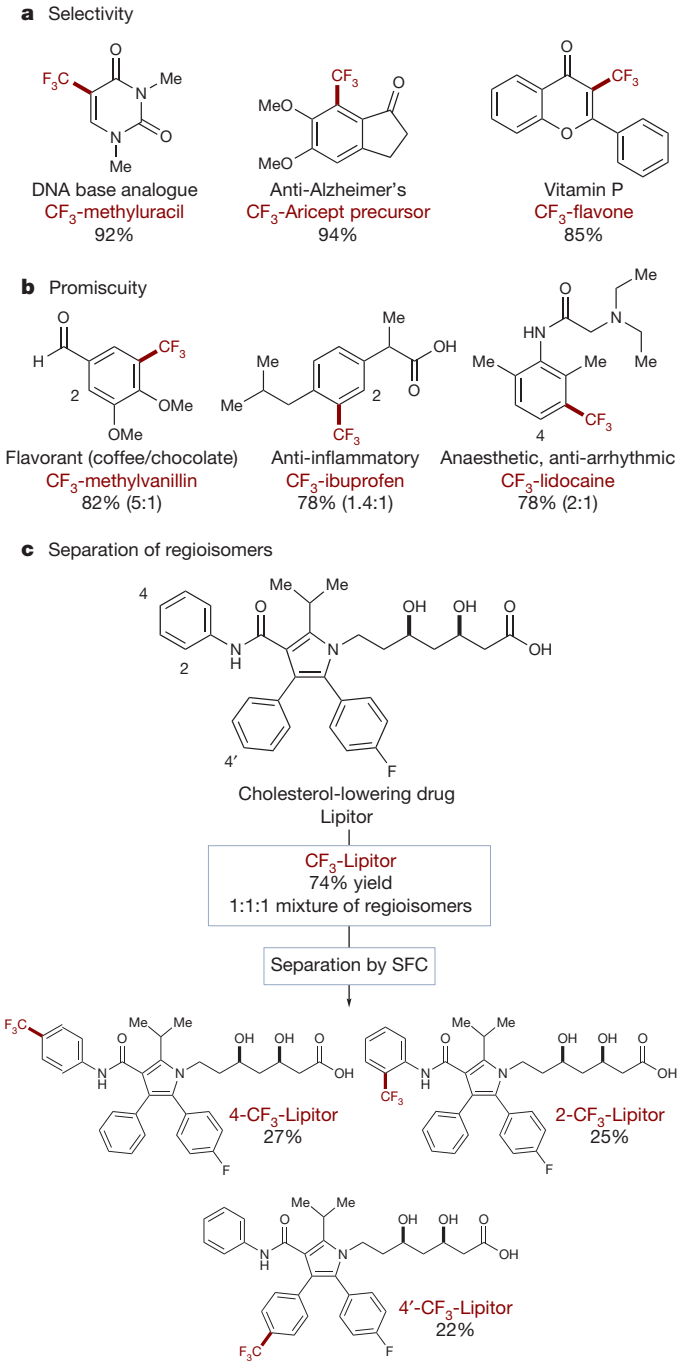


Figure 4 | Direct trifluoromethylation of biologically active molecules. Subjecting common medicinal agents and other biologically active molecules to our standard photoredox protocol enables direct CF_3 installation selectively at metabolically susceptible positions of some molecules (**a**). Alternatively, C–H functionalization of more metabolically stable medicines occurs non-selectively, allowing for rapid access to drug analogues (**b**). The promiscuous modification of equally reactive π-systems, such as the arenes in Lipitor, may be followed by separation of isomers (for example, via supercritical fluid chromatography, SFC) for rapid screening of biological activity (**c**).

substrates, although their use of peroxides as a radical initiator affords divergent substrate tolerance and reaction efficiency. We expect these two simple protocols to be of broad utility for the synthesis and development of new medicinal agents.

METHODS SUMMARY

General procedure. An oven-dried vial was equipped with photocatalyst (1–2 mol%), dry K_2HPO_4 (3 equiv.), arene or heteroarene (0.5 mmol, 1 equiv.) and MeCN (0.125 M), and degassed by alternating vacuum evacuation and argon backfill at −78 °C. The triflyl chloride (1–4 equiv.) was then added and the solution was stirred at room temperature adjacent to a 26-W compact fluorescent light bulb. After 24 h, the reaction mixture was purified by column chromatography to furnish the desired CF_3 product. Full experimental details and characterization data (^{1}H, ^{13}C, ^{19}F NMR, infrared spectroscopy, high-resolution mass spectrometry) for all new compounds are included in Supplementary Information.

Received 12 August; accepted 17 October 2011.

1. Li, J. J. & Johnson, D. S. (eds) *Modern Drug Synthesis* (Wiley, 2010).
2. Montellano, P. R. O. (ed.) *Cytochrome P450: Structure, Mechanism, and Biochemistry* (Springer, 2005).
3. Filler, R., Kobayashi, Y. & Yagupolskii, L. M. (eds) *Organofluorine Compounds in Medicinal Chemistry and Biomedical Applications* (Elsevier, 1993).
4. Muller, K., Faeh, C. & Diederich, F. Fluorine in pharmaceuticals: looking beyond intuition. *Science* **317,** 1881–1886 (2007).
5. Purser, S., Moore, P. R., Swallow, S. & Gouverneur, V. Fluorine in medicinal chemistry. *Chem. Soc. Rev.* **37,** 320–330 (2008).
6. Hagmann, W. K. The many roles for fluorine in medicinal chemistry. *J. Med. Chem.* **51,** 4359–4369 (2008).
7. Tomashenko, O. A. & Grushin, V. V. Aromatic trifluoromethylation with metal complexes. *Chem. Rev.* **111,** 4475–4521 (2011).
8. Furuya, T., Kamlet, A. S. & Ritter, T. Catalysis for fluorination and trifluoromethylation. *Nature* **473,** 470–477 (2011).
9. Oishi, M., Kondo, H. & Amii, H. Aromatic trifluoromethylation catalytic in copper. *Chem. Commun.* 1909–1911 (2009).
10. Cho, E. J. *et al.* The palladium-catalyzed trifluoromethylation of aryl chlorides. *Science* **328,** 1679–1681 (2010).
11. Wang, X., Truesdale, L. & Yu, J. Q. Pd (II)-catalyzed ortho-trifluoromethylation of arenes using TFA as a promoter. *J. Am. Chem. Soc.* **132,** 3648–3649 (2010).
12. Xu, J. *et al.* Copper-catalyzed trifluoromethylation of aryl boronic acids using a CF_3^+ reagent. *Chem. Commun.* **47,** 4300–4302 (2011).
13. Nicewicz, D. A. & MacMillan, D. W. C. Merging photoredox catalysis with organocatalysis: the direct asymmetric alkylation of aldehydes. *Science* **322,** 77–80 (2008).
14. Yoon, T. P., Ischay, M. A. & Du, J. Visible light photocatalysis as a greener approach to photochemical synthesis. *Nature Chem.* **2,** 527–532 (2010).
15. Narayanam, J. M. R. & Stephenson, C. R. J. Visible light photoredox catalysis: applications in organic synthesis. *Chem. Soc. Rev.* **40,** 102–113 (2011).
16. Juris, A. *et al.* Ru(II) polypyridine complexes: photophysics, photochemistry, electrochemistry, and chemiluminescence. *Coord. Chem. Rev.* **84,** 85–277 (1988).
17. Nagib, D. A., Scott, M. E. & MacMillan, D. W. C. Enantioselective α-trifluoromethylation of aldehydes via photoredox organocatalysis. *J. Am. Chem. Soc.* **131,** 10875–10877 (2009).
18. Andrieux, C. P., Gelis, L., Medebielle, M., Pinson, J. & Saveant, J.-M. Outer-sphere dissociative electron transfer to organic molecules: a source of radicals or carbanions? Direct and indirect electrochemistry of perfluoroalkyl bromides and iodides. *J. Am. Chem. Soc.* **112,** 3509–3520 (1990).
19. Skarda, V. *et al.* Luminescent metal complexes. Part 3. Electrochemical potentials of ground and excited states of ring-substituted 2,2′-bipyridyl and 1,10-phenanthroline tris-complexes of ruthenium. *J. Chem. Soc. Perkin Trans. 2* 1309–1311 (1984).
20. Heaton, C. A., Miller, A. K. & Powell, R. L. Predicting the reactivity of fluorinated compounds with copper using semi-empirical calculations. *J. Fluor. Chem.* **107,** 1–3 (2001).
21. Heaton, C. A. & Powell, R. L. Introduction of perfluoroalkyl groups — a new approach. *J. Fluor. Chem.* **45,** 86 (1989).
22. Kamigata, N., Fukushima, T. & Yoshida, M. Reactions of perfluoroalkanesulfonyl chlorides with aromatic compounds catalyzed by a ruthenium (II) complex. *Chem. Lett.* **19,** 649–650 (1990).
23. Kamigata, N., Ohtsuka, T., Fukushima, T., Yoshida, M. & Shimizu, T. Direct perfluoroalkylation of aromatic and heteroaromatic compounds with perfluoroalkanesulfonyl chlorides catalysed by a ruthenium(II) phosphine complex. *J. Chem. Soc. Perkin Trans. 1* 1339–1346 (1994).
24. Dolbier, W. Fluorinated free radicals. *Top. Curr. Chem.* **192,** 97–163 (1997).
25. Langlois, B. R., Laurent, E. & Roidot, N. Trifluoromethylation of aromatic compounds with sodium trifluoromethanesulfinate under oxidative conditions. *Tetrahedron Lett.* **32,** 7525–7528 (1991).
26. Wiehn, M. S., Vinogradova, E. V. & Togni, A. Electrophilic trifluoromethylation of arenes and N-heteroarenes using hypervalent iodine reagents. *J. Fluor. Chem.* **131,** 951–957 (2010).
27. Bahtia, K. & Schuler, R. H. Oxidation of hydroxycyclohexadienyl radical by metal ions. *J. Phys. Chem.* **78,** 2335–2338 (1974).
28. Dexter, D. L., Wolberg, W. H., Ansfield, F. J., Helson, L. & Heidelberger, C. The clinical pharmacology of 5-trifluoromethyl-2′-deoxyuridine. *Cancer Res.* **32,** 247–253 (1972).
29. Roth, B. D. The discovery and development of atorvastatin, a potent novel hypolipidemic agent. *Prog. Med. Chem.* **40,** 1–22 (2002).
30. Ji, Y. *et al.* Innate C–H trifluoromethylation of heterocycles. *Proc. Natl Acad. Sci. USA* **108,** 14411–14415 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

©2011 Macmillan Publishers Limited. All rights reserved

Acknowledgements Financial support was provided by the NIH General Medical Sciences (R01 01 GM093213-01) and gifts from Merck, Amgen, Abbott and Bristol-Myers Squibb. We thank C. Kraml and N. Byrne of Lotus Separations LLC for their development of preparatory supercritical fluid chromatography (SFC) methods and for the separation of all three CF_3-Lipitor analogues.

Author Contributions D.A.N. performed and analysed experiments. D.A.N. and D.W.C.M. designed experiments to develop this reaction and probe its utility, and also prepared this manuscript. Correspondence and requests for materials should be addressed to D.W.C.M. (dmacmill@princeton.edu).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.W.C.M. (dmacmill@princeton.edu).

©2011 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature10626

Combined obliquity and precession pacing of late Pleistocene deglaciations

Peter Huybers[1]

Milankovitch[1] proposed that Earth resides in an interglacial state when its spin axis both tilts to a high obliquity and precesses to align the Northern Hemisphere summer with Earth's nearest approach to the Sun. This general concept has been elaborated into hypotheses that precession[2], obliquity[3,4] or combinations of both[5–8] could pace deglaciations during the late Pleistocene[9,10]. Earlier tests have shown that obliquity paces the late Pleistocene glacial cycles[4,11] but have been inconclusive with regard to precession, whose shorter period of about 20,000 years makes phasing more sensitive to timing errors[4,11,12]. No quantitative test has provided firm evidence for a dual effect. Here I show that both obliquity and precession pace late Pleistocene glacial cycles. Deficiencies in time control that have long stymied efforts to establish orbital effects on deglaciation are overcome using a new statistical test that focuses on maxima in orbital forcing. The results are fully consistent with Milankovitch's proposal but also admit the possibility that long Southern Hemisphere summers contribute to deglaciation.

During the late Pleistocene—roughly over the past million years—Northern Hemisphere continental ice has alternately covered much of northern North America and Fennoscandia and then retreated to today's relatively ice-free conditions at intervals of approximately 100,000 years (100 kyr). The cause of these massive shifts in climate remains unclear not for lack of models, of which there are now over thirty[2–10,13,14], but for want of means to choose among them. Previous statistical tests have demonstrated that obliquity paces the ~100-kyr glacial cycles[4,11], helping narrow the list of viable mechanisms, but have been inconclusive with respect to precession (that is, $P > 0.05$) because of small sample sizes and uncertain timing[4,11,12].

Whether precession influences the ~100-kyr glacial–interglacial cycles is not obvious. Precession alters diurnal average insolation intensity by as much as 30 W m^{-2} on a given day of the year, suggesting a powerful forcing, and its signature clearly appears in proxy records of temperature and ice volume at ~20-kyr periods[2]. However, its insolation anomalies are exactly counterbalanced across the seasons so that annual insolation at any latitude is independent of precession[15]. Furthermore, proxies of early-Pleistocene glaciation show strong obliquity and little precession variability, indicating that precession had negligible influence during this next-most-recent epoch of glaciation[16], though see ref. 17 for another view.

Here, I test whether anomalously large combinations of precession and obliquity forcing combine to determine when deglaciations occurred during the late Pleistocene. The test involves three steps. The first is to estimate the timing of terminations, for which I use a composite δ^{18}O record whose chronology is derived by linearly interpolating age with depth between the last deglaciation and radiometrically dated geomagnetic reversals[11]. Timescale uncertainty over the past million years is estimated by running a stochastic sediment accumulation rate model[11] that also accounts for uncertainties in the alignment of features between δ^{18}O stratigraphies, decompaction of sediment, transport times of δ^{18}O within the ocean, and geomagnetic reversal ages. The age of the Matuyama–Brunhes geomagnetic reversal was earlier assumed to be known to within ±2 kyr (one standard deviation, 1 s.d.)[11], but to account for uncertainty in the ^{40}K decay constant[18,19], it is now represented as occurring at 780 ± 8 kyr (1 s.d.). Terminations are identified by local maxima in the time rate-of-change of the δ^{18}O record that exceed a value of 0.095‰ per kyr, giving the usual termination features[20] except that termination 3 contains two parts that are labelled 3a and 3b (Fig. 1a). (Thresholds ranging between 0.07‰ and 0.17‰ per kyr would give different numbers of terminations but give similarly significant results.) The average uncertainty in the age of the 12 identified termination features is 8 kyr (1 s.d.), with older ages generally being more uncertain.

The second step is to define an insolation forcing function, of which there are many varieties[1,2,16]. For present purposes only the relative shape of the forcing function is needed, and a generic and broadly representative formulation[5] can be adopted:

$$\mathcal{F}_t = \alpha^{1/2} e_t \sin(\omega_t - \phi) + (1-\alpha)^{1/2} \varepsilon_t \quad (1)$$

Here e represents eccentricity, ω is the angle from vernal equinox to perihelion, ε is obliquity, subscript 't' indicates time, and ϕ and α are adjustable parameters that respectively control the phase of precession and the relative contributions from precession and obliquity. Both $e_t\sin(\omega_t - \phi)$ and ε_t are normalized to zero-mean and unit variance such that $\mathcal{F}_t$ also has unit variance. Milankovitch[1], along with many subsequent authors[5–7,9,10,16], called upon anomalies in incoming solar radiation during the Northern Hemisphere summer to determine whether the Northern Hemisphere is glaciated. Increased insolation intensity during Northern Hemisphere summer results from greater obliquity and a phase of precession that brings Earth closer to the Sun during that season, and can be represented by setting $\phi = 0^\circ$ and $\alpha = 0.5$. The resulting structure of $\mathcal{F}_t$ (Fig. 1b) shares more than 99% of its variance in common with both Milankovitch's caloric summer half-year insolation[1] at 65° N and summer energy[16] at 65° N when using a threshold of 350 W m^{-2}, providing a suitable representation of the hypothesis that Northern Hemisphere summer insolation controls glaciation.

In the third and final step, forcing maxima in $\mathcal{F}_t$ that correspond most closely in time with each termination are compared against forcing maxima not associated with terminations (Fig. 1b). In particular, the median value of the non-termination maxima is subtracted from the median value of the termination maxima, yielding $\delta m = 0.89$. Unlike the Rayleigh's R statistic (relied upon for previous tests of orbital influence upon deglaciation[4,11,12,21]), timing errors do not affect δm unless they cause the wrong forcing cycles to be identified. Median values are also less sensitive to timing errors because outliers generally have no effect.

The significance of δm is assessed within the context of a null hypothesis H_0, that termination timing is independent of $\mathcal{F}_t$, and an alternative hypothesis H_1, that terminations tend to occur when the maxima in $\mathcal{F}_t$ are anomalously large. For purposes of comparison with earlier work[4], a modified random walk representing ice-volume variability is adopted for the null hypothesis:

$$v_t = v_{t-1} + \eta_t \quad \text{and if } v_t \geq h_t \text{ then terminate} \quad (2)$$

Ice volume v_t accumulates by a random increment η_t during each 1-kyr time step, until a threshold h_t is passed, and the termination of all ice is

[1]Harvard University, Department of Earth and Planetary Sciences, 20 Oxford Street, Cambridge, Massachusetts 02138, USA.

©2011 Macmillan Publishers Limited. All rights reserved

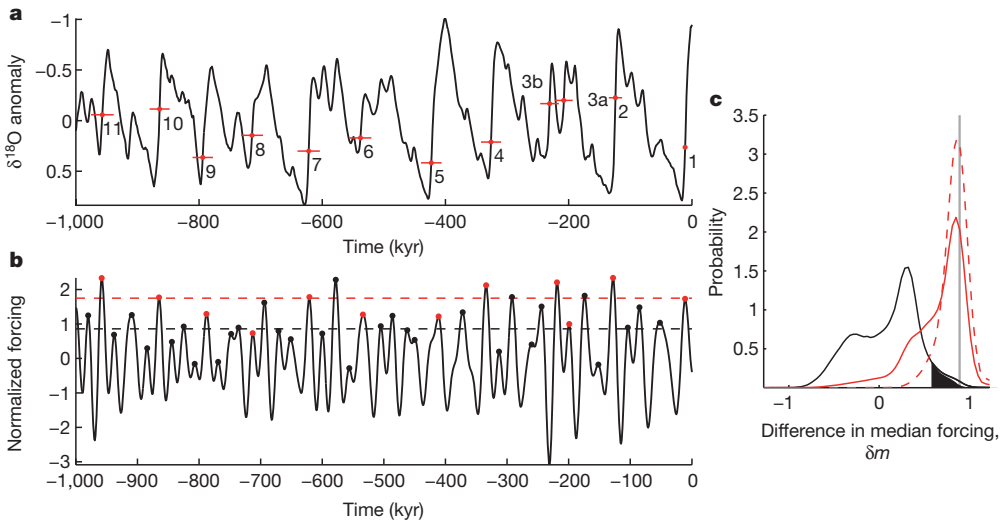


Figure 1 | Test of the Milankovitch hypothesis. **a**, Anomalies in δ^{18}O on a depth-derived timescale[11]. Terminations are identified when rates of change exceed 0.95‰ per kyr and are pinpointed at their maximum rate of change (red dots, with 2 s.d. time-uncertainty indicated by the horizontal bars). **b**, Orbital forcing containing equal amounts of obliquity and climatic precession variability and the contribution of climatic precession peaking during the Northern Hemisphere summer solstice. The forcing maximum nearest each termination is indicated (red dots), and the median across all such maxima (red dashed line) is 0.89 normalized units greater than the median of the maxima not associated with terminations (black dots and dashed line). **c**, The difference in the medians (vertical grey bar) is highly significant ($P = 0.002$) as judged against the null hypothesis (H_0, black with the 5% rejection region filled). The difference in medians is also consistent with the alternative hypothesis (H_1, red lines), whether timing errors are ignored (dashed) or accounted for (solid).

prescribed to occur over the ensuing 10 kyr. Defining $h_t = 90$ units, randomly initializing ice volume between 0 and 90 units, and drawing η_t from a normal distribution with unit mean and two units of standard deviation gives glacial cycles with a period of 100 ± 20 kyr, consistent with observations[4]. Terminations are identified from realizations of v_t in the same manner as for the δ^{18}O record, except that v_t is first smoothed using an eleven-point running average, and rates of ice volume change must be less than -5 units per kyr. Given these random realizations of termination times, a value of δm is computed from $\mathcal{F}_t$, and the process repeated 10^5 times to build up a distribution for H_0 (Fig. 1c). This null-distribution of δm is shifted towards positive values because of a tendency for large-amplitude forcing cycles to have a longer period, making H_0 harder to reject than if the nonparametric Wilcoxon rank sum test[22] had instead been used.

H_1 is similarly derived from equation (2), but after modifying the threshold condition to be sensitive to insolation forcing, $h_t = 110 - 25\mathcal{F}_t$, and prescribing η_t to have unit variance, which again yields glacial cycles of 100 ± 20 kyr. Under H_1, terminations tend to trigger when both ice volume and $\mathcal{F}_t$ are large. Termination times calculated according to H_1 are then perturbed using realizations from the stochastic sediment accumulation model consistent with the errors estimated for the depth-derived ages (see Fig. 1a).

The statistical power of the test is measured as the probability of rejecting H_0 at the $P = 0.05$ significance level if H_1 is correct. In the absence of timing errors, the statistical power would be 0.95, but even when timing errors are accounted for, it is still 0.67. Although the power of the test depends upon the form and parameters chosen for the null and alternative models, these results indicate that, unlike for previous tests[4,11,12], the combined influence of obliquity and precession will be identifiable, if present. Indeed, the observed value of $\delta m = 0.89$ is highly significant as judged against H_0 ($P = 0.002$) and is near the maximum likelihood value of H_1 (Fig. 1c), even after accounting for errors in timing. Therefore, it is possible to reject the null hypothesis of no orbital influence with a high degree of confidence, and the results are consistent with precession and obliquity together exerting control upon deglaciation.

To explore the sensitivity of these results to different formulations of the orbital forcing, I repeated the test for $\mathcal{F}_t$ computed using all phases of precession and all mixtures of obliquity and precession (Fig. 2). Such combinations reproduce the variability expressed by other measures of insolation forcing, including the caloric summer half-year[1], summer energy[16] (as long as thresholds are not set near peak intensity), and insolation at any particular latitude and on any particular day of the year[23]. The resulting contour of significance levels shows highly significant results ($P < 0.01$) in the vicinity of perihelion aligning with the Northern Hemisphere summer solstice and equal contributions from obliquity and precession, though significant results ($P < 0.05$) are also found for perihelion occurring a month earlier or several months later and forcing mixtures having α between 0.3 and 0.9. This shows that any formulation of $\mathcal{F}_t$ in general agreement with Milankovitch's hypothesis would also give a positive result.

To specifically inquire into the ability to reject the null hypothesis of obliquity-only pacing of the glacial cycles, I adopt a null hypothesis wherein the threshold in equation (2) varies according to

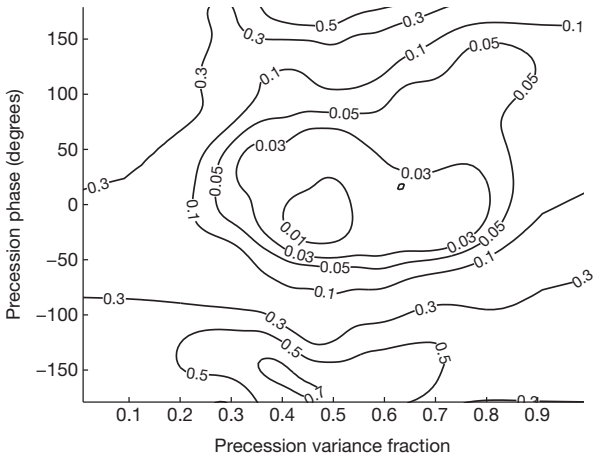


Figure 2 | Sensitivity of the test to choice of orbital forcing. Most orbital forcing curves can be represented as a linear mixture of obliquity and climatic precession (x axis, giving the fraction of precession variance), where climatic precession can take on any phase (y axis, where 0° indicates greatest forcing during the Northern Hemisphere summer solstice). The test's P-value is contoured after applying it to each forcing function. Values less than 0.05 are significant, and the results are consistent with Milankovitch's hypothesis. Forcing functions with a negative obliquity contribution (not shown) are never significant. A two-dimensional smoother was applied to the logarithm of the P-values to make the plot easier to interpret while preserving small values.

©2011 Macmillan Publishers Limited. All rights reserved

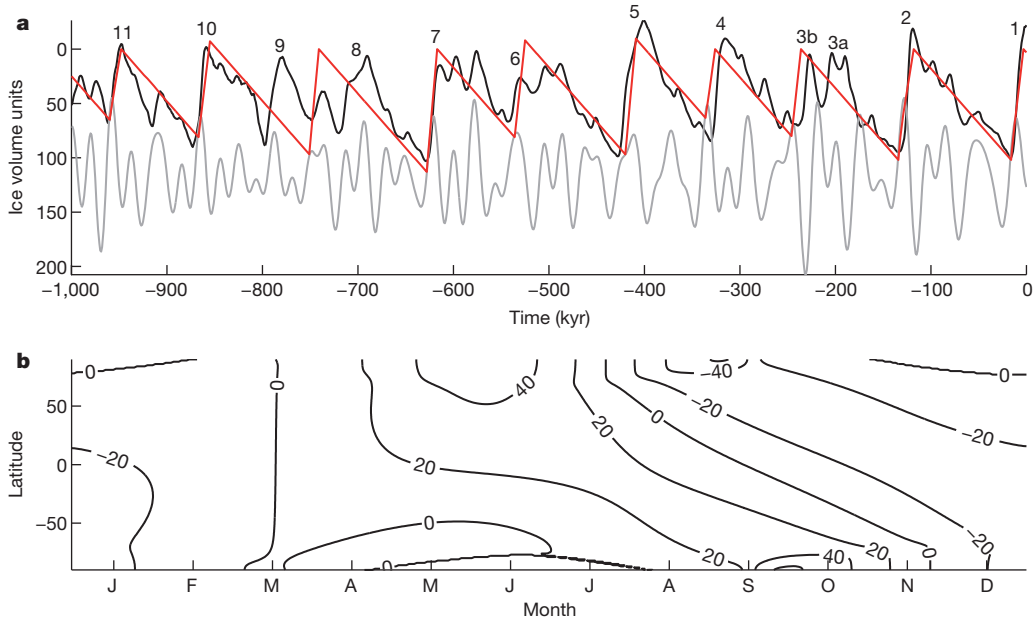


Figure 3 | Orbital forcing of glacial cycles. a, A deterministic model of the glacial cycles (red line) using the same threshold condition (grey line) for the alternate hypothesis. For reference, a scaled version of the composite δ^{18}O record is also shown (black line). With the exceptions of terminations 3a, 8 and 9 the timing and amplitude of the glacial cycles are generally reproduced. **b**, The insolation anomaly (in W m^{-2}) that drives deglaciations. Anomalies are computed as the average insolation across the forcing maxima associated with terminations (see Fig. 1) relative to the average insolation over the past million years. Note that anomalies are shown relative to day of the year, not solar longitude[30].

$h_t = 105 - 15\varepsilon_t$, a formulation consistent with ref. 4 wherein obliquity paces the timing of deglaciation. This formulation shifts the null distribution towards higher values but nonetheless shows $\delta m = 0.89$ to be significant ($P = 0.02$), a result that also holds when any other combination of mean and amplitude is specified for the threshold. Likewise, adoption of a precession-only null hypothesis invariably leads to $\delta m = 0.89$ being significant ($P < 0.01$), regardless of specifications for the mean and amplitude of the threshold and the phase of precession. (The obliquity-only null is harder to reject than the precession-only null because it leads to selection of forcing maxima in $\mathcal{F}_t$ that tend to have the positive phases of both obliquity and precession aligned.) These results show that glacial hypotheses calling singly on precession[2] or obliquity[3,4], or not accounting for either[13,14], should be rejected in favour of combined obliquity and precession pacing of late Pleistocene glacial cycles[1,5–8,10].

I also repeated the test after updating the depth-derived age model using a compilation of available radiometric constraints upon the age of the last four terminations[24]. The most significant revision is to shift the ages of terminations 3a and 3b to align with older precession maxima, which increases the mean of the deglacial forcing maxima but has no effect on the median. Therefore, the revised ages do not change the test results, reflecting the test's overall robustness to timing errors. The data and code used to generate each of the above results are available in the Supplementary Information.

The above statistical results can be succinctly illustrated by replacing the stochastic variable in the alternative hypothesis with its mean value and initializing ice volume to be 25 units at −1,000 kyr. The resulting deterministic time-history of glaciation gives a cross-correlation of 0.73 with the negative of the δ^{18}O record, which is very high relative to other simple model fits given the small number of adjustable parameters[25] (Fig. 3a). The model result reproduces the basic sawtooth structure of the late-Pleistocene glacial cycles over the past million years in terms of timing and amplitude, except for terminations 8 and 3b, and illustrates how obliquity, precession and the growth of ice volume can combine to control late-Pleistocene glacial cycles. Another implication is that the climate will tend to be driven out of glaciation sooner when eccentricity—and hence the amplitude of precession—is larger. Because interglacial values of δ^{18}O are less variable than glacial values, this relationship implies that the amplitude of the ~100-kyr glacial cycles are smaller when eccentricity is larger, consistent with the findings reported in ref. 21.

Combined orbital pacing is also consistent with earlier findings that the intervals between successive deglaciations cluster into 80-kyr or 120-kyr periods[4,11], indicative of two or three obliquity cycles. Precession, with its ~20-kyr period, achieves a maximum during nearly every interval of above-average obliquity, given the ~40-kyr period of obliquity, but the reverse does not hold. Therefore, precession will tend to influence the precise timing of a deglaciation within an obliquity cycle, but obliquity will more fundamentally govern the interval between deglaciations. Whether precession also influences the precise timing of terminations during the early Pleistocene, when deglaciations occur more nearly every 40 kyr, remains an open question[16,17]. Perhaps terminations then occurred nearly every time precession maxima coincided with above-average values of obliquity; such a scenario can be obtained by setting the threshold in equation (2) to lower values.

Ice sheets tend to collapse in response to unusually large maxima in insolation forcing that result from the coincidence of high obliquity and alignment of perihelion with Northern Hemisphere summer solstice, consistent with the models hypothesized by Milankovitch[1] and others[5–10]. During these forcing maxima, summer insolation is as much as 40 W m^{-2} greater at high northern latitudes (Fig. 3b). However, this consistency is not exclusive of all other orbital contributions to deglaciation. For instance, when perihelion aligns with the Northern Hemisphere summer solstice, aphelion occurs during the Southern Hemisphere summer, causing the length of the Southern Hemisphere summer to be longer (Fig. 3b) and, possibly, increasing the escape of CO_2 from the Southern Ocean into the atmosphere[26–29]. The climate system is thoroughly interconnected across temporal and spatial scales, and, just as neither obliquity nor precession act in isolation, no one region should be expected to exert exclusive influence upon deglaciation.

Received 14 August; accepted 11 October 2011.

1. Milankovitch, M. *Kanon der Erdbestrahlung und seine Anwendung auf das Eiszeitenproblem* 1–503 (Royal Serbian Academy, 1941).
2. Hays, J., Imbrie, J. & Shackleton, N. Variations in the Earth's orbit: pacemaker of the ice ages. *Science* **194,** 1121–1132 (1976).

©2011 Macmillan Publishers Limited. All rights reserved

3. Liu, H. Phase modulation effect of the Rubincam insolation variations. *Theor. Appl. Climatol.* **61,** 217–229 (1998).
4. Huybers, P. & Wunsch, C. Obliquity pacing of the late Pleistocene glacial terminations. *Nature* **434,** 491–494 (2005).
5. Imbrie, J. & Imbrie, J. Modeling the climatic response to orbital variations. *Science* **207,** 943–953 (1980).
6. Paillard, D. The timing of Pleistocene glaciations from a simple multiple-state climate model. *Nature* **391,** 378–381 (1998).
7. Berger, A., Li, X. & Loutre, M. Modelling northern hemisphere ice volume over the last 3Ma. *Quat. Sci. Rev.* **18,** 1–11 (1999).
8. Tziperman, E., Raymo, M., Huybers, P. & Wunsch, C. Consequences of pacing the Pleistocene 100 kyr ice ages by nonlinear phase locking to Milankovitch forcing. *Paleoceanography* **21,** PA4206 (2006).
9. Berger, A. Milankovitch theory and climate. *Rev. Geophys.* **26,** 624–657 (1988).
10. Saltzman, B. *Dynamical Paleoclimatology: Generalized Theory of Global Climate Change* (Academic Press, 2002).
11. Huybers, P. Glacial variability over the last two million years: an extended depth-derived agemodel, continuous obliquity pacing, and the Pleistocene progression. *Quat. Sci. Rev.* **26,** 37–55 (2007).
12. Kawamura, K. *et al.* Northern Hemisphere forcing of climatic cycles in Antarctica over the past 360,000 years. *Nature* **448,** 912–916 (2007).
13. Ghil, M. Cryothermodynamics: the chaotic dynamics of paleoclimate. *Physica D* **77,** 130–159 (1994).
14. Wunsch, C. The spectral description of climate change including the 100ky energy. *Clim. Dyn.* **20,** 353–363 (2003).
15. Herschel, J. On the astronomical causes which may influence geological phenomena. *Trans. Geol. Soc. Lond.* **3,** 293–300 (1832).
16. Huybers, P. Early Pleistocene glacial cycles and the integrated summer insolation forcing. *Science* **313,** 508–511 (2006).
17. Raymo, M., Lisiecki, L. & Nisancioglu, K. Plio-Pleistocene ice volume, Antarctic climate, and the global $\delta^{18}O$ record. *Science* **313,** (2006).
18. Min, K., Mundil, R., Renne, P. & Ludwig, K. A test for systematic errors in 40Ar/39Ar geochronology through comparison with U/Pb analysis of a 1.1-Ga rhyolite. *Geochim. Cosmochim. Acta* **64,** 73–98 (2000).
19. Kuiper, K. *et al.* Synchronizing rock clocks of Earth history. *Science* **320,** 500–504 (2008).
20. Broecker, W. In *Milankovitch and Climate* (eds Berger, A. *et al.*) Part 2, 687–698 (D. Riedel, 1984).
21. Lisiecki, L. Links between eccentricity forcing and the 100,000-year glacial cycle. *Nature Geosci.* **3,** 349–352 (2010).
22. Gibbons, J. *Nonparametric Methods for Quantitative Analysis* 1–487 (American Sciences Press, 1985).
23. Imbrie, J. *et al.* On the structure and origin of major glaciation cycles. I: Linear responses to Milankovitch forcing. *Paleoceanography* **7,** 701–738 (1992).
24. Siddall, M., Chappell, J. & Potter, E. In *Developments in Quaternary Science* Vol. 7, 75–92 (Elsevier, 2007).
25. Roe, G. & Allen, M. A comparison of competing explanations for the 100,000-yr ice age cycle. *Geophys. Res. Lett.* **26,** 2259–2262 (1999).
26. Broecker, W. & Henderson, G. The sequence of events surrounding Termination II and their implications for the cause of glacial-interglacial CO_2 changes. *Paleoceanography* **13,** 352–364 (1998).
27. Schulz, K. & Zeebe, R. Pleistocene glacial terminations triggered by synchronous changes in Southern and Northern Hemisphere insolation: the insolation canon hypothesis. *Earth Planet. Sci. Lett.* **249,** 326–336 (2006).
28. Huybers, P. & Denton, G. Antarctic temperature at orbital time scales controlled by local summer duration. *Nature Geosci.* **1,** 787–792 (2008).
29. Timmermann, A., Timm, O., Stott, L. & Menviel, L. The roles of CO_2 and orbital forcing in driving Southern Hemispheric temperature variations during the last 21,000 years. *J. Clim.* **22,** 1626–1640 (2009).
30. Joussaume, S. & Braconnot, P. Sensitivity of paleoclimate simulation results to season definitions. *J. Geophys. Res.* **102,** 1943–1956 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work benefited from comments by G. Gebbie, L. Lisiecki, A. Stine and C. Wunsch.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to the author (phuybers@fas.harvard.edu).

©2011 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature10589

Frictional ageing from interfacial bonding and the origins of rate and state friction

Qunyang Li[1]†, Terry E. Tullis[2], David Goldsby[2] & Robert W. Carpick[1]

Earthquakes have long been recognized as being the result of stick–slip frictional instabilities[1,2]. Over the past few decades, laboratory studies of rock friction have elucidated many aspects of tectonic fault zone processes and earthquake phenomena[3–5]. Typically, the static friction of rocks grows logarithmically with time when they are held in stationary contact[6], but the mechanism responsible for this strengthening is not understood. This time-dependent increase of frictional strength, or frictional ageing, is one manifestation of the 'evolution effect' in rate and state friction theory[5]. A prevailing view is that the time dependence of rock friction results from increases in contact area caused by creep of contacting asperities[7,8]. Here we present the results of atomic force microscopy experiments that instead show that frictional ageing arises from the formation of interfacial chemical bonds, and the large magnitude of ageing at the nanometre scale is quantitatively consistent with what is required to explain observations in macroscopic rock friction experiments. The relative magnitude of the evolution effect compared with that of the 'direct effect'—the dependence of friction on instantaneous changes in slip velocity—determine whether unstable slip, leading to earthquakes, is possible[9,10]. Understanding the mechanism underlying the evolution effect would enable us to formulate physically based frictional constitutive laws, rather than the current empirically based 'laws'[11,12], allowing more confident extrapolation to natural faults.

Frictional ageing has been attributed to increases in contact area caused by asperity creep (contact 'quantity') as well as to time-dependent strengthening of bonding at asperity contacts (contact 'quality'). Evidence exists to support both interpretations. By measuring the amount of light transmitted across rough Lucite plastic surfaces in contact, and rough soda-lime glass surfaces in contact, Dieterich and Kilgore[8] observed gradual increases in the sizes of illuminated microscopic contacts over time. Experiments on a silicate rock (quartzite) have demonstrated that frictional ageing is suppressed by drying the samples and conducting friction experiments in a water-free environment[7]. Because creep is inhibited in experiments on silicate minerals like quartz at high temperature in the absence of water (that is, in the absence of hydrolytic weakening[13]), these observations are consistent with the hypothesis that static friction increases with changes in contact area caused by asperity creep[14]. However, frictional ageing may also result from strengthening of chemical bonding at the interface over time. Chemical bonding could be enhanced by time-dependent desorption of contamination films[15] that may include water, or through chemically assisted mechanisms (such as hydrogen bonding or siloxane bridging)[16–18] that can be aided by water and also affected by the contact stresses.

Devising experiments to understand the contribution of each mechanism to ageing is exceedingly difficult, because the buried frictional interface is not readily accessible and involves myriad microscopic asperity contacts with a range of sizes down to the nanoscale[19,20]. Here, we investigate the ageing behaviour of a single nanoscale contact to isolate the physical origins of the ageing process and provide new perspectives on the empirical rate- and state-dependent friction laws.

We conducted single-asperity slide–hold–slide (SA-SHS) friction experiments between nanoscopically sharp oxidized silicon tips and oxidized silicon wafers using an atomic force microscope (AFM); these tests are analogous to similar slide–hold–slide tests on rocks[3,6]. The use of amorphous SiO_2 (silica) to investigate the behaviour of rocks composed of silicate minerals is reasonable, because silica glass and rocks behave similarly in friction experiments on rough surfaces[21].

Figure 1 shows an example of SA-SHS measurements performed at 60% relative humidity (RH). Similar to slide–hold–slide experiments on rocks[3,6], the static friction force is larger than the steady-state sliding friction force by an amount ΔF, called the friction drop, that grows linearly with the logarithm of the hold time (inset to Fig. 1). Similar trends are obtained at other humidities. The linear dependence of ΔF on the logarithm of hold time is consistent with observations from friction experiments on macroscopic rocks and other materials[6,16], but the magnitude observed here is far larger. The ratio between ΔF and the steady-state friction F_{ss} ($\Delta F/F_{ss}$), which we call the 'relative friction drop', ranges from 0.5 to 5 over the range of humidity studied for 100-s hold times, compared to less than 0.05 for a 100-s hold time observed for macroscopic quartz rock samples[16]. This unusually large ageing effect is reproducible for all the SA-SHS experiments, with more than five distinct silica–silica pairs tested.

The much larger ageing of the relative friction drop at the nanoscale than at the macroscale suggests that frictional ageing may be a length-scale-dependent phenomenon influenced by the multi-asperity

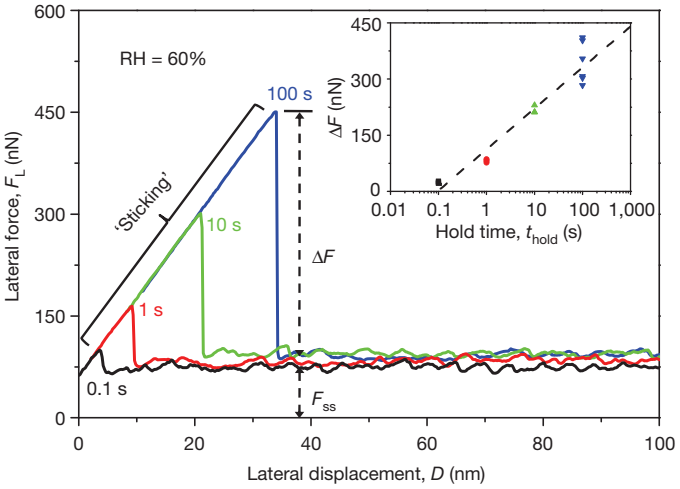


Figure 1 | Lateral force versus nominal lateral displacement data for typical SA-SHS tests after stationary holds at 60% RH. Upon lateral displacement, the tip sticks to the substrate, resulting in linear, elastic lateral loading of the AFM cantilever (Supplementary Figs 1 and 2). When the lateral force exceeds static friction, the tip slips forward, indicated by abrupt drops in lateral force (ΔF), followed by subsequent sliding at the steady-state friction force (F_{ss}). In the inset ΔF varies linearly with the logarithm of hold time. The dotted line is a linear fit of the averaged values.

[1]Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [2]Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA. †Present address: School of Aerospace, Tsinghua University, Beijing 100084, China.

©2011 Macmillan Publishers Limited. All rights reserved

character of rough surfaces. To test this hypothesis, we construct a simplified mechanics model to predict friction as a function of nominal contact size for a fixed number of discrete asperities per unit area (Fig. 2). Comparing a small (19 asperities) and large (1,459 asperities) nominal contact area, with the same individual asperity (i) relative friction drop ($\Delta F^i/F^i_{ss}=1$), the total macroscopic relative friction drop ($\Delta F^{total}/F^{total}_{ss}$) values are very different: 0.81 for the smaller nominal contact and 0.15 for the larger nominal contact. The behaviour can be understood as follows. Owing to elastic interactions between asperities through the substrate, the asperities do not experience the same shear stress and thus do not slide simultaneously; the shear stress rises from the contact centre outward, until slip occurs. For sufficiently small nominal contact areas, asperities slide nearly at the same time, so macroscopic ageing is similar to individual asperity ageing. However, for larger nominal contacts, only a small fraction of asperities experience the peak forces at any given time: those in the outermost annulus have slid and thus relaxed the peak stress to the much lower steady-state sliding value, thus providing little extra resistance, whereas those in the centre have not yet been stressed to the peak shear stress, thus providing little contribution to the measured friction force. Therefore, only a small internal annulus of asperities exhibit the peak shear stress, and hence $\Delta F^{total}/F^{total}_{ss}$ of a nominally micrometre-sized contact interface that has nanoscale asperities will be less than 10% of the corresponding values for the single asperities.

Although the model is highly idealized, it demonstrates a general behaviour: larger nominal contacts will generally have a much lower macroscopic ageing effect than that measured for individual asperities within the nominal contact area regardless of the specific mechanism(s) for ageing. This can account for the different magnitudes of frictional ageing seen in rock friction experiments on rough surfaces[6,16] and in the single-asperity experiments. This scale dependence is somewhat analogous to that observed for adhesion: small objects adhere to surfaces far more strongly (relative to their weight) than macroscopic objects, owing to the numerous asperities on rough surfaces of large objects[22].

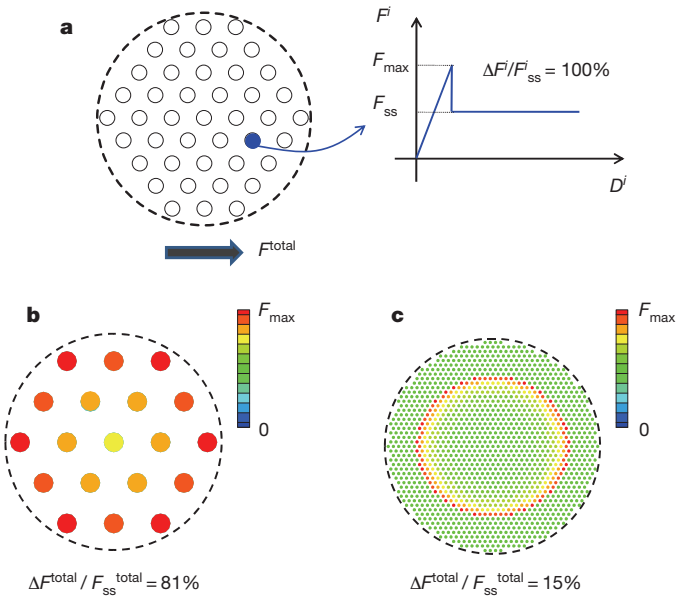


Figure 2 | The quasi-static discrete asperity model. a, An interface between a rigid top surface and an elastic half space with a circular nominal contact area encompassing identical close-packed circular asperities at fixed spacing (small circles), all with maximum static friction force F_{max} (with $\Delta F^i/F^i_{ss}=1$). Upon lateral loading by F^{total}, asperities are uniformly elastically strained up to their maximum static friction, whereupon they slip (see Supplementary Information). The shear force distributions at the peak frictional force for 19 (**b**) and 1,459 (**c**) asperities result in macroscopic relative static friction ($\Delta F^{total}/F^{total}_{ss}$) that is high (0.81) for **a** but much lower (0.15) for **c**.

Although the unusually large ageing effect at the single-asperity level revealed by our experiments is required to explain the frictional ageing observed in friction experiments on rough macroscopic rock surfaces, the results presented thus far do not identify the underlying cause of this large ageing effect. In the present SA-SHS experiments, we found that ageing can depend critically on the chemical state of the AFM tip (Fig. 3). We found that frictional ageing was substantially but reversibly suppressed when the AFM tip was pulled out of contact and exposed to the humid environment before undergoing an SA-SHS test; ageing could be revitalized by sliding the tip across the substrate for some cumulative run-in sliding distance (and by not breaking contact immediately before an SA-SHS test). While the tip is out of contact, the tip is susceptible to adsorption of gaseous species, including organics (well known to reduce the surface energy of silica surfaces)[23], whereas during scanning of the tip on the substrate, such adsorption is suppressed and adsorbates can be removed. This clearly indicates that plastic deformation of the tip is not involved in the frictional ageing observed in the SA-SHS tests, given that lifting up and sliding the tip would not influence plastic deformation of the tip in such a reversible manner.

The surfaces of oxidized silicon samples cleaned with Piranha solution (5:1 concentrated sulphuric acid to 30% hydrogen peroxide solution) are highly hydroxylated[24,25]. In a humid environment, hydrogen bonding can arise between two hydroxylated silica surfaces[17,18,25], which, together with van der Waals and capillary interactions, will increase the overall interfacial adhesion and friction. Furthermore, at raised temperatures the silanol (Si–OH) groups from opposing sides of contacting surfaces can react to form stronger siloxane (Si–O–Si) bonds[17,25]. At room temperature (typically 23 °C), with applied load and sufficient time, the reaction may also proceed[17], but the kinetics have yet to be studied in detail. Capillary interactions are unlikely to be a dominant mechanism because of the persistence of ageing at humidities below 1% RH (see Supplementary Fig. 5). Therefore, we hypothesize that frictional ageing is caused by hydrogen or siloxane bond formation across the interface.

To test this hypothesis, we performed another set of SA-SHS tests by sliding a silica tip over two chemically inert surfaces: hydrogen-terminated diamond, and graphite. The diamond sample is grown by chemical vapour deposition (typical grain size ~1 μm) and subsequently exposed to atomic hydrogen using an established method[26] that produces a highly uniform, nearly oxygen-free, monohydride surface before the friction experiment. The graphite surface was prepared by freshly cleaving a bulk highly ordered pyrolytic graphite sample. Interfacial bonding is not expected between silica and either

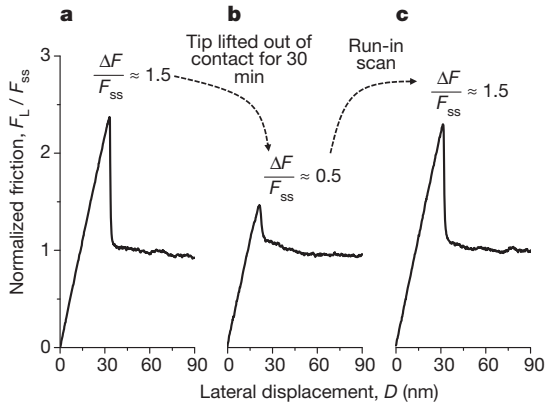


Figure 3 | Normalized friction–displacement curves from three sequential SA-SHS tests. a, Typical, strong ageing is seen immediately after a first run-in scan (cumulative displacement of ~1 mm); $\Delta F/F_{ss}=1.5$. **b**, Ageing is substantially suppressed after 30 min of exposure of the tip to the humid environment, out of contact with the sample, followed immediately by the SA-SHS test shown; $\Delta F/F_{ss}=0.5$. **c**, Large ageing returned to the original value after sliding for another run-in scan; $\Delta F/F_{ss}=1.5$. All tests used a 100-s hold time at 40% RH, with a negative lateral hold force (see Supplementary Information).

©2011 Macmillan Publishers Limited. All rights reserved

H-terminated diamond or graphite. The normalized friction versus displacement plots from SA-SHS tests (100-s hold time) are shown in Fig. 4 and compared with data from the oxidized silicon surface. In contrast to the significant frictional ageing (static friction increases by a factor of six compared to the steady-state friction) for the silica–silica surface, the silica–diamond interface and the silica–graphite interface show little to no ageing. Contact stresses are estimated to be about 0.2 GPa, 0.4 GPa and 1.5 GPa for silica–silica, silica–graphite and silica–diamond contacts respectively (see Supplementary Fig. 8). If plastic creep occurs for the silica–silica contact, we would also expect it in the tip for all of the systems, leading to a similarly high degree of ageing (perhaps even more for the silica–diamond contact where the contact stresses were largest). However, given the indentation hardness of amorphous silica of approximately 12 GPa (ref. 27), we would not expect plastic creep in the tip for any of these material pairs at the normal stresses involved. Given both this expectation and the results shown in Figs 1, 3 and 4, our experiments clearly demonstrate that the frictional ageing effect in SA-SHS tests does not primarily arise from plastic creep, but from changes in bonding arising from surface chemical processes.

In conclusion, we find that frictional strength for AFM tips of silica sliding on a silica substrate increases linearly as a function of the logarithm of contact time, in the same fashion as the frictional strengths of silicate minerals in macroscopic friction experiments. However, the magnitude of the ageing effect for the nanoscale single-asperity contacts is much larger than for macroscopic multi-asperity rock friction experiments. We demonstrate that this difference should be expected given the scales and single- versus multi-asperity geometries of the two cases regardless of the specific mechanisms for ageing. At the nanometre scale and at relatively low contact stresses, frictional ageing cannot be explained by increases in contact area due to plastic creep. Rather, it is associated with changes in chemical bonding as demonstrated by the dependence of ageing on the identity of the samples that are in contact with the AFM tips, and the reversible suppression of ageing by removing the tip from otherwise continuous contact with the sample. This reduction of ageing due to exposure of the interface to ambient conditions may contribute partially to the low frictional ageing observed in macroscopic rock friction experiments. In such experiments, both contacting surfaces are usually rough and the local asperities can be exposed to water or other contaminants intermittently during sliding. However, this effect alone is insufficient to account for the very large difference in ageing we observe; variations in relative frictional ageing before and after breaking the contact within our experiment is typically between 0.5–5, whereas the difference between our results and macroscopic experiments is two orders of magnitude.

The magnitude of ageing from these chemical-bonding effects is large enough to account for the evolution effect in rate and state friction without a time-dependent increase in contact area. Whether the primary cause of frictional ageing in macroscopic friction experiments is an increase in the quantity and strength of chemical bonding or an increase in contact area due to plastic creep is unresolved by our data. However, we have demonstrated that chemical bonding is a viable explanation, although we have not determined the specific chemical mechanism that leads to such a slow, logarithmic ageing. The logarithmic behaviour demonstrates that the rate of reaction is slowing down as a function of time. Further experimental and theoretical studies are needed to address the mechanism(s) that account for this.

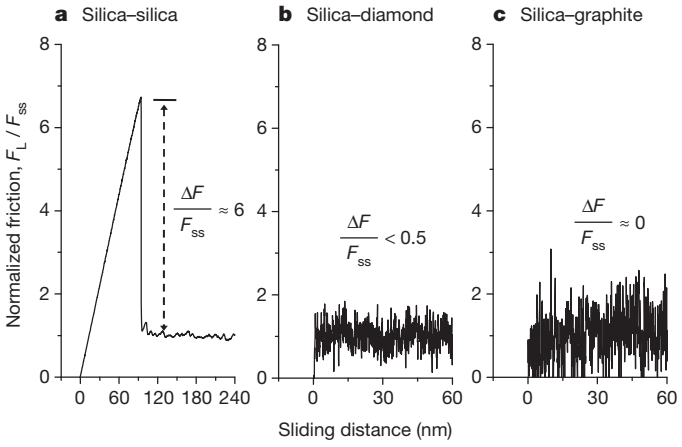


Figure 4 | Three SA-SHS tests between a silica tip and three different surfaces. **a**, Silica–silica; **b**, silica–hydrogen-terminated diamond; **c**, silica–graphite. The tests were all performed at 60% RH for a 100-s hold time. The normal load in each case is maintained at approximately 1 nN. The lateral forces during stationary hold are negative (see Supplementary Information).

METHODS SUMMARY

Experiments were conducted by sliding silica AFM tips over silica surfaces using friction force microscopy (see Methods and Supplementary Fig. 1). The probes and substrate are processed by heating silicon probes and Si(100) wafers in pure oxygen to form oxides 100 nm and 400 nm thick respectively. The substrates were cleaned with Piranha solution, rinsed with de-ionized water, dried under pure nitrogen flow, then rapidly inserted into the AFM. The substrates have a typical root-mean-square roughness of ~0.2 nm within a 0.5 μm × 0.5 μm area, according to the AFM. Normal and lateral force constants of the probes were calibrated using Sader's method[28] and a diamagnetic lateral force calibrator[29] respectively. All measurements are carried out in a sealed chamber with RH varied from <1% to 80%. The system was equilibrated for at least 4 h after each humidity change to reduce thermal positional drift, which is particularly important when long hold times are used. The remaining drift may contribute to the increased scatter seen at longer hold times. This increased scatter may also be due to higher sensitivity of the interfacial strength to any inhomogeneities at the interface. RH was monitored by a hygrometer (uncertainty ±1.5% RH). A typical friction experiment is performed by first making contact between the tip and substrate, then sliding for an initial run-in distance. Then, without breaking contact, a slide–hold–slide procedure is used in which the tip is slid for a prescribed distance (typically ~400 nm), then held stationary for a time t_{hold}, and then lateral displacement is resumed. Constant normal load is imposed by a feedback loop. Four different hold times, in an alternating sequence of 0.1 s, 100 s, 1 s and 10 s, are used. The lateral force during the stationary hold was maintained at either a positive or negative value (see Methods, Supplementary Figs 2, 6 and 7 and associated discussion).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 May; accepted 21 September 2011.
Published online 30 November 2011.

1. Brace, W. F. & Byerlee, J. D. Stick-slip as a mechanism for earthquakes. *Science* **153,** 990–992 (1966).
2. Scholz, C. H. Earthquakes and friction laws. *Nature* **391,** 37–42 (1998).
3. Marone, C. Laboratory-derived friction laws and their application to seismic faulting. *Annu. Rev. Earth Planet. Sci.* **26,** 643–696 (1998).
4. Scholz, C. H. *The Mechanics of Earthquakes and Faulting* 2nd edn (Cambridge University Press, 2002).
5. Tullis, T. E. Rock friction constitutive behavior from laboratory experiments and its implications for an earthquake prediction field monitoring program. *Pure Appl. Geophys.* **126,** 555–588 (1988).
6. Dieterich, J. H. Time-dependent friction in rocks. *J. Geophys. Res.* **77,** 3690–3697 (1972).
7. Dieterich, J. H. & Conrad, G. Effect of humidity on time-dependent and velocity-dependent friction in rocks. *J. Geophys. Res.* **89** (B6), 4196–4202 (1984).
8. Dieterich, J. H. & Kilgore, B. D. Direct observation of frictional contacts—new insights for state-dependent properties. *Pure Appl. Geophys.* **143,** 283–302 (1994).
9. Ranjith, K. & Rice, J. R. Stability of quasi-static slip in a single degree of freedom elastic system with rate and state dependent friction. *J. Mech. Phys. Solids* **47,** 1207–1218 (1999).
10. Rice, J. R., Lapusta, N. & Ranjith, K. Rate and state dependent friction and the stability of sliding between elastically deformable solids. *J. Mech. Phys. Solids* **49,** 1865–1898 (2001).
11. Dieterich, J. H. Modeling of rock friction. 1. Experimental results and constitutive equations. *J. Geophys. Res.* **84** (B5), 2161–2168 (1979).
12. Ruina, A. Slip instability and state variable friction laws. *J. Geophys. Res.* **88** (B12), 10359–10370 (1983).

©2011 Macmillan Publishers Limited. All rights reserved

13. Griggs, D. Hydrolytic weakening of quartz and other silicates. *Geophys. J. R. Astron. Soc.* **14,** 19–31 (1967).
14. Scholz, C. H. & Engelder, J. T. The role of asperity indentation and ploughing in rock friction–I: Asperity creep and stick-slip. *Int. J. Rock Mech. Min. Sci. Geomech. Abstr.* **13,** 149–154 (1976).
15. Hirth, J. & Rice, J. On the thermodynamics of adsorption at interfaces as it influences decohesion. *Metall. Mater. Trans. A* **11,** 1501–1511 (1980).
16. Frye, K. M. & Marone, C. Effect of humidity on granular friction at room temperature. *J. Geophys. Res.* **107,** 2309, doi:10.1029/2001JB000654 (2002).
17. Michalske, T. A. & Fuller, E. R. Closure and repropagation of healed cracks in silicate glass. *J. Am. Ceram. Soc.* **68,** 586–590 (1985).
18. Stavrinidis, B. & Holloway, D. G. Crack healing in glass. *Phys. Chem. Glasses* **24,** 19–25 (1983).
19. Ben-David, O., Rubinstein, S. M. & Fineberg, J. Slip-stick and the evolution of frictional strength. *Nature* **463,** 76–79 (2010).
20. Bowden, F. P. & Tabor, D. *The Friction and Lubrication of Solids* Ch. I (Oxford University Press, 1958).
21. Weeks, J. D., Beeler, N. M. & Tullis, T. E. Frictional behavior: glass is like a rock. *Eos* **72,** (Fall Meeting Abs. Supp.) 457–458 (1991).
22. Kendall, K. *Molecular Adhesion and its Applications: The Sticky Universe.* (Springer, 2001).
23. Roche, A. *et al.* in *Ultra Clean Processing of Silicon Surfaces 2000* (eds Heyns, M., Mertens, P. & Meuris, M.) 76–77, 111–114 (Scitec, 2001).
24. Iler, R. K. *Chemistry of Silica—Solubility, Polymerization, Colloid and Surface Properties and Biochemistry* Ch. 1 (John Wiley & Sons, 1979).
25. Ploessl, A. & Krauter, G. Wafer direct bonding: tailoring adhesion between brittle materials. *Mater. Sci. Eng. Rep.* **25,** 1–88 (1999).
26. Thoms, B. D., Owens, M. S., Butler, J. E. & Spiro, C. Production and characterization of smooth, hydrogen-terminated diamond c(100). *Appl. Phys. Lett.* **65,** 2957–2959 (1994).
27. Oliver, W. C. & Pharr, G. M. An improved technique for determining hardness and elastic-modulus using load and displacement sensing indentation experiments. *J. Mater. Res.* **7,** 1564–1583 (1992).
28. Sader, J. E., Chon, J. W. M. & Mulvaney, P. Calibration of rectangular atomic force microscope cantilevers. *Rev. Sci. Instrum.* **70,** 3967–3969 (1999).
29. Li, Q., Kim, K. S. & Rydberg, A. Lateral force calibration of an atomic force microscope with a diamagnetic levitation spring system. *Rev. Sci. Instrum.* **77,** 065105 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M.O. Robbins, I. Szlufarska and Y. Liu for discussions. We acknowledge support from the National Science Foundation under awards EAR0810088 and EAR0810192.

Author Contributions Q.L. performed experiments and obtained the data, and analysed the data with input from all other authors. All four authors wrote this manuscript together. D.G. prepared some of the tips used for the experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.W.C. (carpick@seas.upenn.edu).

©2011 Macmillan Publishers Limited. All rights reserved

METHODS

Friction force microscopy. Friction force microscopy was conducted using an RHK UHV350 AFM. A schematic of the setup of a friction force microscope is shown in Supplementary Fig. 1. The microscope consists of a sharp tip in contact with the surface of the sample. The tip is integrated with the end of a microfabricated cantilever that deforms in response to normal and lateral (friction) forces arising from the contacting interface. Deformation of the cantilever is measured with an optical system consisting of a laser and a quadrant photo detector. For frictional sliding, the back end of the cantilever is displaced relative to the sample along the indicated scan direction. Because of the extremely small tip radius, the lateral size of the contacting interface is usually a few to tens of nanometres, which provides an ideal single-asperity contact. During a friction experiment, the relative lateral displacement between the sample and the holder is increased at a constant rate and the resulting lateral force is measured while the normal load is maintained constant via an active feedback loop.

Normal and lateral force constants of the probes were calibrated using Sader's method[28] and a diamagnetic lateral force calibrator[29] respectively. All measurements were carried out in a sealed chamber with RH varied by bubbling pure dry nitrogen vapour from a liquid nitrogen dewar through water (or pure nitrogen, for measurements quoted at $<$1% RH). The system was equilibrated for at least 4 h after each humidity change. RH was monitored by a hygrometer (uncertainty $\pm$1.5% RH).

Sample preparation. The probes and substrate were processed by heating silicon probes (CSC37 Mikromasch) and Si(100) wafers in pure oxygen for 150 min and 300 min, respectively, to form oxides 100 nm and 400 nm thick (determined by ellipsometry), respectively. The substrates were cleaned with Piranha solution, rinsed with de-ionized water, dried under pure nitrogen flow, then rapidly inserted into the AFM. The substrates have a typical root-mean-square roughness of about 0.2 nm within 0.5 μm $\times$ 0.5 μm area as measured by an AFM.

The hydrogen-terminated diamond sample was grown by chemical vapour deposition (typical grain size $\sim$1 μm) and subsequently exposed to atomic hydrogen using an established method[26]. The graphite surface was prepared by freshly cleaving bulk highly ordered pyrolytic graphite.

SA-SHS experiments. A typical friction experiment is performed by first making contact between the tip and substrate, then sliding for an initial run-in distance (typically $\sim$1 mm, in successive 1-μm back-and-forth lateral displacements). Then, without breaking contact, a slide–hold–slide procedure was used in which the tip is slid for a prescribed distance (typically $\sim$400 nm), then held stationary for a time t_{hold}, and then lateral displacement is resumed. Constant normal load is imposed by a feedback loop. Four different hold times, in an alternating sequence of 0.1 s, 100 s, 1 s and 10 s, were used.

Two protocols are used as shown by the schematics in Supplementary Fig. 2. Supplementary Fig. 2a shows the positive lateral hold force protocol. The tip is slid for an initial distance (a[1], a[2]), then held stationary for a time t_{hold} without changing the lateral force (a[3]). Sliding is then reinitiated in the same direction by ramping up the lateral force (a[4]). Supplementary Fig. 2b shows the negative lateral hold force protocol. The tip is slid for an initial distance (b[1]), and then held stationary for a time t_{hold} without changing the lateral force (b[2]). Sliding is then reinitiated in the opposite direction by ramping up the lateral force (b[3], b[4]). The lateral force during the hold is 'positive' and usually very close to the steady-state friction value for the positive lateral hold force protocol, and is 'negative' and very close to the steady-state friction level for the negative lateral hold force protocol. The results shown in the main text are obtained using the positive hold force protocol; the results from the negative hold force protocol are given in Supplementary Figs 6 and 7 for comparison.

©2011 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature10689

Acute vision in the giant Cambrian predator *Anomalocaris* and the origin of compound eyes

John R. Paterson[1], Diego C. García-Bellido[2], Michael S. Y. Lee[3,4], Glenn A. Brock[5], James B. Jago[3,6] & Gregory D. Edgecombe[7]

Until recently[1], intricate details of the optical design of non-biomineralized arthropod eyes remained elusive in Cambrian Burgess-Shale-type deposits, despite exceptional preservation of soft-part anatomy in such Konservat-Lagerstätten[2,3]. The structure and development of ommatidia in arthropod compound eyes support a single origin some time before the latest common ancestor of crown-group arthropods[4], but the appearance of compound eyes in the arthropod stem group has been poorly constrained in the absence of adequate fossils. Here we report 2–3-cm paired eyes from the early Cambrian (approximately 515 million years old) Emu Bay Shale of South Australia, assigned to the Cambrian apex predator *Anomalocaris*. Their preserved visual surfaces are composed of at least 16,000 hexagonally packed ommatidial lenses (in a single eye), rivalling the most acute compound eyes in modern arthropods. The specimens show two distinct taphonomic modes, preserved as iron oxide (after pyrite) and calcium phosphate, demonstrating that disparate styles of early diagenetic mineralization can replicate the same type of extracellular tissue (that is, cuticle) within a single Burgess-Shale-type deposit. These fossils also provide compelling evidence for the arthropod affinities of anomalocaridids, push the origin of compound eyes deeper down the arthropod stem lineage, and indicate that the compound eye evolved before such features as a hardened exoskeleton. The inferred acuity of the anomalocaridid eye is consistent with other evidence that these animals were highly mobile visual predators in the water column[5,6]. The existence of large, macrophagous nektonic predators possessing sharp vision—such as *Anomalocaris*—within the early Cambrian ecosystem probably helped to accelerate the escalatory 'arms race' that began over half a billion years ago[7,8].

Anomalocaridids are broadly acknowledged to be the top predators in Cambrian and Ordovician ecosystems[6,9]. Predatory habits are inferred on the basis of their large size (body length more than 91.5 cm, ref. 9), robust spinose frontal appendages, mouth with a dentate inner margin[10], size and form of the midgut glands[11], predation damage to associated biota[10,12] and interpretation of coprolites as having an anomalocaridid origin[6,13]. The streamlined profile of the body, inferred function of the lateral body flaps ('swim flaps'[14]) in locomotion, and the large tripartite tail fan preserved in a few taxa indicate strong swimming capabilities[5,15,16]. The large, stalked eyes of anomalocaridids, emerging from the dorsolateral side of the head, provide additional evidence for prey detection and tracking consistent with predatory habits. Although the eyes are preserved in several anomalocaridid genera[13,15,17,18], they have until now been known solely from their outlines. The inference that they were probably compound eyes[18] has been based on size, shape and the phylogenetic placement of anomalocaridids in the arthropod stem group[18–20] rather than on direct evidence of surface detail (such as preserved lenses).

Exceptionally preserved eyes from the early Cambrian (Series 2, Stage 4) Emu Bay Shale[21] at Buck Quarry, Kangaroo Island, South Australia, are much larger, and have very different morphology, than those of another arthropod documented from this locality[1]. The visual surface is pyriform (pear-shaped) rather than hemispherical, and the ommatidia are more than five times as numerous, yet much smaller in maximum size, with no evidence of a distinct 'bright zone'. The specimens are registered in the collections of the South Australian Museum. SAM P45920a,b (part and counterpart) is a pair of eyes (Fig. 1) that are in contact with each other at their proximal ends, their long axes diverging from each other at 45 degrees. Both eyes are of the same size and morphology, and a chance association can be ruled out; they are identified as the left and right eyes of a single individual; although the eyes in most other anomalocaridids are widely spaced[10,13,17,18], the preserved orientation of the eyes in SAM P45920 is similar to that of *Amplectobelua symbrachiata* from China (figure 15A in ref. 15) and probably results from lateral or oblique compression of the head. They are each composed of a pyriform visual surface with a very large number of small, hexagonal ommatidial lenses. In the more complete of the pair (Fig. 1d), the long axis of the visual surface is 21.8 mm long, its maximum width 12.2 mm (including a fracture, Fig. 1b). Part of this eye is overprinted across its width by a different structure of undetermined nature ('us' in Fig. 1b, d). It is preserved in positive relief on the part, with its lenses concave. The other eye has a maximum width of 12.7 mm; it is preserved in negative relief on the part, and its lenses are convex. In the counterpart, the concavity/convexity of the visual surface and lenses is reversed. We interpret the sediment in the proximal portion of this eye as having in-filled the collapsed cavity, and the cuticle is now replicated by iron oxide. Lenses range from ~70–110 μm in diameter, with variability throughout the visual field attributed to surface irregularities. The lenses are consistently arranged with regular hexagonal packing relative to their neighbours (Fig. 1c, e). Extrapolating from the regions with the best preserved lenses across the area of the entire visual surface of the eye in Fig. 1d indicates that an estimated 16,700 lenses are present on one side of the eye. The proximo-distal axis of the visual field is established on the basis of the preservation of the eye stalk on the left eye, with a sharp delineation from the visual surface (Fig. 1b, d). The eye stalk lacks obvious surface sculpture and is more strongly developed along one side of the visual surface than the other. A second specimen of a single eye, P46330a,b (Fig. 2b and Supplementary Fig. 1a–e), is even larger, 16.9 mm wide, and has patches of well-preserved lenses, but is too incomplete to measure its length.

Scanning electron microscopic energy dispersive spectrometry (SEM-EDS) analysis of SAM P45920a shows that the visual surface is composed of iron oxide with detectable traces of sulphur (Fig. 2a), probably indicating limonite after pyrite. Limonite pseudomorphs of microcrystalline pyrite are often found concentrated beneath and between non-trilobite arthropod cuticle in the Emu Bay Shale (figure 7 in ref. 22), as are peripheral haloes of iron oxide around a variety of fossils, suggesting that pyritization was a common preservation mode

[1]Division of Earth Sciences, School of Environmental and Rural Science, University of New England, Armidale, New South Wales 2351, Australia. [2]Departamento de Geología Sedimentaria y Cambio Ambiental, Instituto de Geociencias (CSIC-UCM), José Antonio Novais 2, Madrid 28040, Spain. [3]South Australian Museum, North Terrace, Adelaide, South Australia 5000, Australia. [4]School of Earth and Environmental Sciences, University of Adelaide, South Australia 5005, Australia. [5]Department of Biological Sciences, Macquarie University, New South Wales 2109, Australia. [6]School of Natural and Built Environments, University of South Australia, Mawson Lakes, South Australia 5095, Australia. [7]Department of Palaeontology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK.

©2011 Macmillan Publishers Limited. All rights reserved

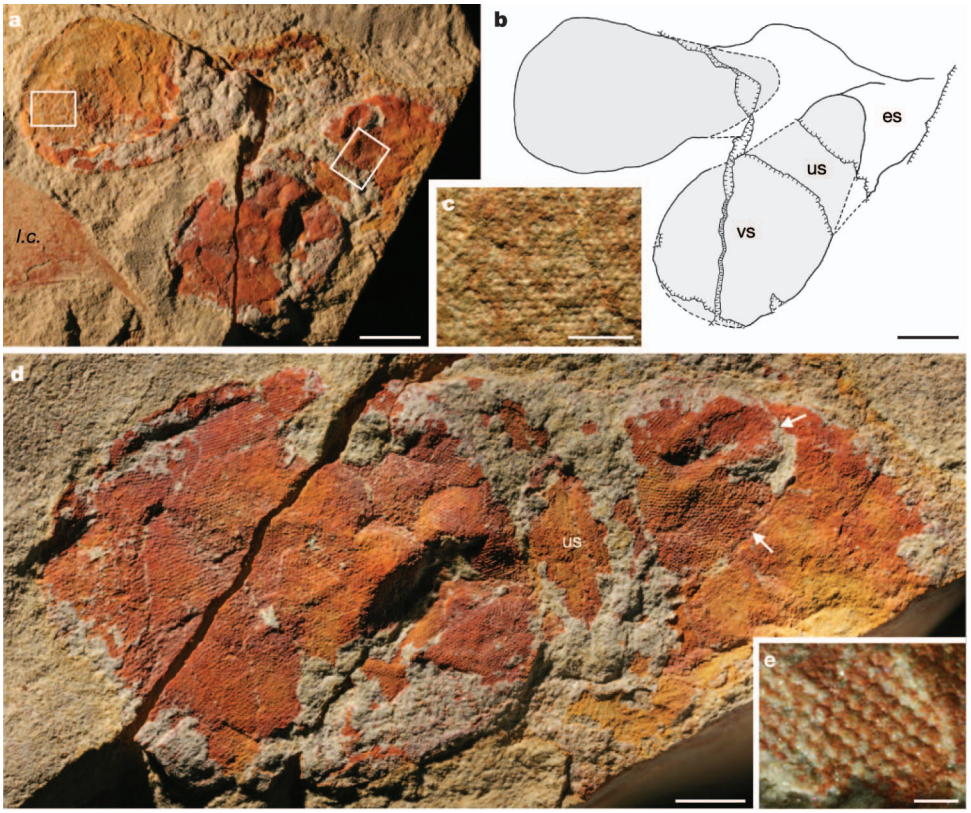


Figure 1 | *Anomalocaris* eyes from the Emu Bay Shale. **a**–**d**, Eye pair, SAM P45920a, level 10.4 m. **a**, **b**, Overview and camera lucida drawing. Scale bars, 5 mm. Grey fill in **b** represents visual surface, the proximal part in the upper eye extrapolated from the lower eye. **c**, Detail of ommatidial lenses located by horizontal white box in **a**. Scale bar, 1 mm. **d**, More complete eye, showing transition between visual surface and eye stalk (white arrows). Scale bar, 2 mm. **e**, Detail of ommatidial lenses in counterpart SAM P45920b. Scale bar, 0.3 mm. es, eye stalk; *l.c.*, *Isoxys communis*; us, undetermined structure; vs, visual surface. Tilted white box in **a** represents area analysed using SEM-EDS, with elemental maps shown in Fig. 2a.

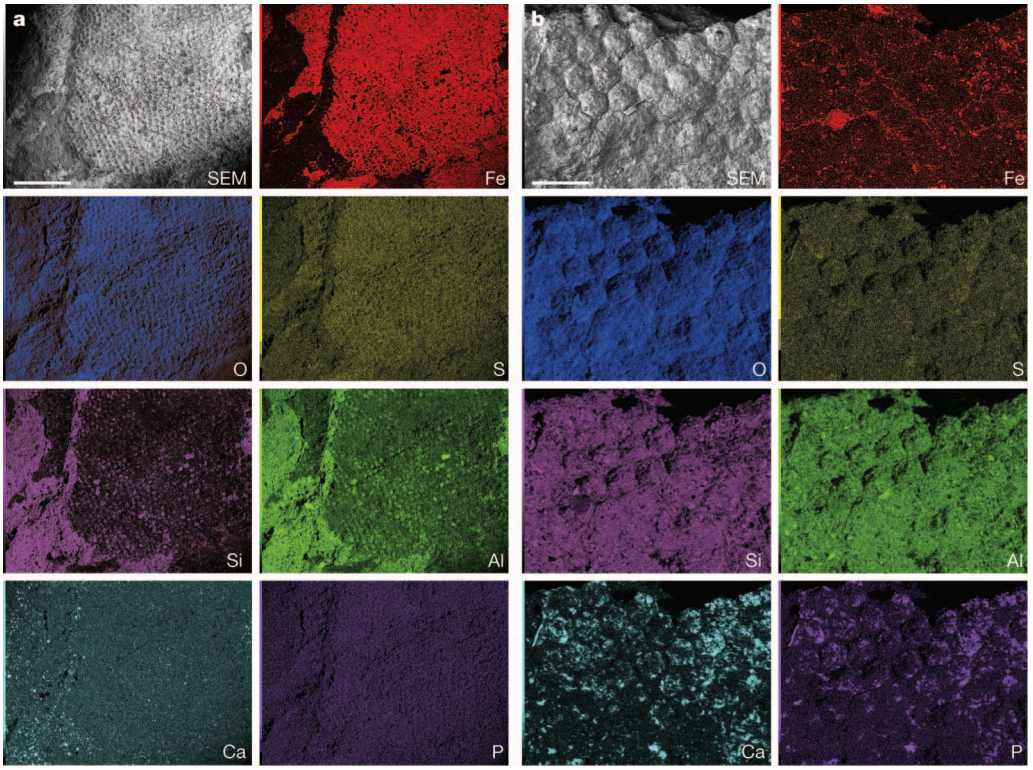


Figure 2 | SEM-EDS analyses of *Anomalocaris* eyes. **a**, SAM P45920a. Scale bar, 1 mm; see Fig. 1a for area analysed. **b**, SAM P46330b. Scale bar, 0.3 mm; see Supplementary Fig. 1e for area analysed. Accelerating voltage of 20 kV. Each map depicts the relative abundance of each element, with brighter colours indicating greater abundance. SEM, backscattered electron image of area analysed. Al, aluminium (green); Ca, calcium (cyan); Fe, iron (red); O, oxygen (dark blue); P, phosphorus (purple); S, sulphur (yellow); Si, silicon (pink). The visual surface of SAM P45920a in **a** contains elevated amounts of iron, oxygen and sulphur, indicative of limonite after pyrite; the matrix (at left of each image) shows high levels of silicon and aluminium, reflecting muscovite and aluminosilicate clay minerals. The lenses of SAM P46330b in **b** contain elevated amounts of calcium and phosphorus, indicative of calcium phosphate.

©2011 Macmillan Publishers Limited. All rights reserved

within this deposit. Moreover, *Anomalocaris* frontal appendages and lateral body flaps (Supplementary Fig. 1f) from the Emu Bay Shale are often preserved as iron oxide, suggesting that pyritization of the entire body was possible, as in the anomalocaridids from the Ordovician Fezouata Biota in Morocco[9] and *Schinderhannes bartelsi* from the Devonian Hunsrück Slate in Germany[19]. Early diagenetic pyritization of soft tissues is also common in the Chengjiang Biota and requires high concentrations of iron in pore waters and low organic content within the sediment to promote reactions between iron minerals and H_2S generated by anaerobic bacterial sulphate reduction[23]. These specific conditions seem to have been present during the deposition of the fossiliferous interval of the lower Emu Bay Shale, the mudstones being iron rich but extremely low in total organic carbon[22]. Interestingly, an SEM-EDS analysis of SAM P46330b (Fig. 2b) reveals that the lenses contain elevated amounts of calcium and phosphorus (and only minor traces of iron), suggesting that phosphatization (rather than pyritization) was the taphonomic pathway in this instance. The role of phosphatization was also important in preserving the eyes of a different arthropod in the Emu Bay Shale[1], demonstrating that disparate styles of early diagenetic mineralization can replicate the same type of extracellular structure (in this case, the cuticle of the visual surface) within a single Burgess-Shale-type deposit. This contrasts with many other Cambrian Konservat-Lagerstätten, wherein recalcitrant tissues (such as cuticle) typically preserve as carbonaceous films[24] and may therefore explain why intricate details of the visual surface (especially in non-biomineralizing arthropods) are rarely preserved[1].

These Emu Bay Shale eyes (Fig. 1 and Supplementary Fig. 1a–e) are identified as belonging to anomalocaridids. Two species of *Anomalocaris* occur in the Emu Bay Shale, *A. briggsi* Nedin, 1995, and *Anomalocaris* sp. nov.[25]. Both taxa are present in Buck Quarry as well as at the locality on the coastline at Big Gully. Frontal appendages and lateral body flaps of *Anomalocaris* are present at the levels (10.4 and 11.3 m) from which the eyes were collected, and are common within a 2-m interval that includes these levels (Supplementary Figs 1f and 2). The large size of the eyes rules out an assignment to all other arthropods known from the Emu Bay Shale, but is consistent in size with the eyes of anomalocaridids known from articulated specimens from other Cambrian Konservat-Lagerstätten (Supplementary Table 1). The shape of the Emu Bay Shale eyes compares closely to that of *Anomalocaris*[13,17], *Amplectobelua*[15], *Laggania*[10] and *Hurdia*[18], all being variably pyriform or ovoid; the length:width ratio (1.8:1) of the present elements falls well within the variation in anomalocaridids as a whole (1.3:1–2.0:1) (Supplementary Table 1). The width of the eye stalk relative to the width of the distal pyriform section of the eye (that bears the visual surface in SAM P45920a) likewise resembles the stalked eyes of articulated anomalocaridids. Considering eye length in articulated specimens of *Anomalocaris* from Canada[17] and China[13] relative to the length of the frontal appendages, the Emu Bay Shale eyes are of an appropriate size relative to the frontal appendages of *A. briggsi*. We thus identify the specimens as *Anomalocaris*, the only anomalocaridid genus known from the Emu Bay Shale.

The number of ommatidia in the *Anomalocaris* eyes would almost certainly have greatly exceeded the count based on the exposed surface of the eye alone. If the flattened surface in the fossils is even partly mirrored on the other side, as suggested by the fact that the three available specimens each have lenses over the entire exposed area of the fossils, but factoring in possible asymmetry to provide cuticular support of the visual surface, the total count could be substantially greater than the observed 16,000+ lenses. If this is indeed the case, few living arthropods have as many ommatidia, and these eyes would certainly have functioned with a high degree of acuity[26]. Assuming that SAM P46330 has the same proportions as the more complete SAM P45920, a length of the visual surface in excess of 3 cm is inferred, with the likelihood of an even greater ommatidial count than in SAM P45920. Throughout the geological history of Arthropoda, compound eyes have rarely exceeded this size; very large Siluro-Devonian pterygotid eurypterids and some Jurassic thylacocephalans represent some of the rare examples with eyes larger than those of *Anomalocaris*.

The discovery of compound eyes in *Anomalocaris* provides compelling support for arthropod affinities[13,18]. Dense, hexagonal packing of ommatidia in compound eyes has been demonstrated to have been unequivocally present in *Schinderhannes bartelsi*, a Devonian species resolved as the immediate sister group to the arthropod crown group[19] (Fig. 3). The eyes of *Schinderhannes* resemble those of *Anomalocaris* in being large, stalked, having an ovoid outline of the visual surface, and a highly elevated number of lenses. The finding that *Anomalocaris*, resolved more basally than *Schinderhannes* in the arthropod stem group[19], possesses the same kind of ommatidial packing as in *Schinderhannes* and crown-group arthropods pushes the origin of compound eyes further down the arthropod stem group. As such, compound eyes evolved earlier than the origin of a hardened tergal exoskeleton and biramous trunk limbs (the latter characters being present in *Schinderhannes* but not anomalocaridids). We infer that the stalked eyes of all Radiodonta[17] (that is, anomalocaridids) are arthropod-type compound eyes. Previous inferences of this character distribution[18] based on gross morphology (size and stalks) now have direct support. The mode of growth of the anomalocaridid eye is presumed to be the same as in other arthropods, in which new elements are added at the margins of the visual field[4]. The next most stem-ward (basal) taxon in the arthropod stem group[18,19], *Opabinia*, also has stalked eyes, but direct observation of ommatidia is needed to ascertain whether they are compound.

The resolving power of compound eyes depends to a large extent on the angle between ommatidia; smaller angles mean denser image sampling and higher acuity. Although exact angles in the *Anomalocaris* eyes cannot be determined owing to extensive compression, upper limits on average angles can be estimated (see Supplementary Methods). Assuming the preserved visual surface does not curve more than 180°, average interommatidial angles would be $<1.4°$; even this conservative estimate suggests greater acuity than most living arthropods[27]. The interommatidial angle and average lens diameter of ~95 μm yields a low 'eye parameter' value of <2, characteristic of diurnal taxa living in

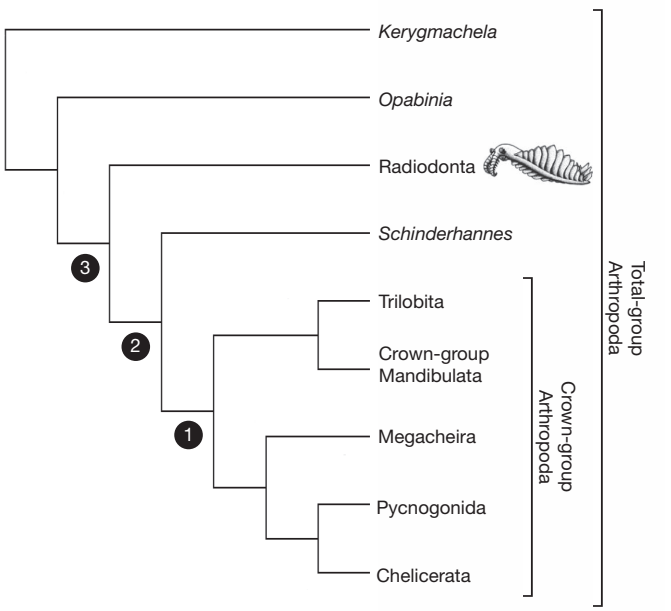


Figure 3 | The early evolution of compound eyes, and the position of anomalocaridids (Radiodonta), in the arthropod stem group. Numbers refer to the inclusiveness of the monophyletic group that can be confidently inferred to possess compound eyes: (1) based on extant taxa alone; (2) based on discovery of *Schinderhannes*[19]; and (3) based on new data herein. Phylogeny after ref. 19.

©2011 Macmillan Publishers Limited. All rights reserved

well-lit marine environments[27], and is consistent with suggestions that anomalocaridids were midwater predators in the photic zone[6].

The very large size of anomalocaridid compound eyes and the visual acuity inferred from the elevated lens number and low interommatidial angles suggest that processing of visual information would have required the optic neuropils and brain to be of comparable complexity to crown-group (that is, modern) arthropods. In the crown group, two optic neuropils are reconstructed in the most recent common ancestor, transmitting to a protocerebrum with a median unpaired neuropil, the central body[28]. The eyes of *Anomalocaris* expand the known diversity of visual adaptations in the early Cambrian: low-resolution organs with $<$100 ommatidia (eodiscoid trilobites), higher-resolution eyes with a distinct bright zone that might have functioned in low light[1], and very large eyes with a uniformly dense visual field adapted to bright environments.

Functional morphological arguments from the structure of raptorial frontal appendages and large, spherical eyes in various Cambrian arthropods indicate that visual predation in the water column was already established as an important component in early Cambrian food webs[29,30]. The large absolute size and huge number of ommatidial lenses in the eyes of *Anomalocaris* confirm its status as a highly visual apex predator[6]. The presence of sophisticated nektonic predators with acute vision, such as anomalocaridids, within Cambrian communities would have placed considerable selective pressures on prey that would have influenced the 'arms race' that began during this important phase in early animal evolution[7,8].

Received 8 September; accepted 28 October 2011.

1. Lee, M. S. Y. *et al.* Modern optics in exceptionally preserved eyes of Early Cambrian arthropods from Australia. *Nature* **474,** 631–634 (2011).
2. Briggs, D. E. G., Erwin, D. H. & Collier, F. J. *The Fossils of the Burgess Shale* (Smithsonian Institution, 1994).
3. Hou, X.-G. *et al. The Cambrian Fossils of Chengjiang, China: The Flowering of Early Animal Life* (Blackwell, 2004).
4. Harzsch, S. & Hafner, G. Evolution of eye development in arthropods: phylogenetic aspects. *Arthropod Struct. Dev.* **35,** 319–340 (2006).
5. Briggs, D. E. G. Giant predators from the Cambrian of China. *Science* **264,** 1283–1284 (1994).
6. Vannier, J. & Chen, J.-Y. Early Cambrian food chain: new evidence from fossil aggregates in the Maotianshan Shale Biota, SW China. *Palaios* **20,** 3–26 (2005).
7. Marshall, C. R. Explaining the Cambrian ''explosion'' of animals. *Annu. Rev. Earth Planet. Sci.* **34,** 355–384 (2006).
8. Bush, A. M. & Bambach, R. K. Paleoecologic megatrends in marine Metazoa. *Annu. Rev. Earth Planet. Sci.* **39,** 241–269 (2011).
9. Van Roy, P. & Briggs, D. E. G. A giant Ordovician anomalocaridid. *Nature* **473,** 510–513 (2011).
10. Whittington, H. B. & Briggs, D. E. G. The largest Cambrian animal, *Anomalocaris*, Burgess Shale, British Columbia. *Phil. Trans. R. Soc. Lond. B* **309,** 569–609 (1985).
11. Butterfield, N. J. *Leanchoilia* guts and the interpretation of three-dimensional structures in Burgess Shale-type fossils. *Paleobiology* **28,** 155–171 (2002).
12. Nedin, C. *Anomalocaris* predation on nonmineralized and mineralized trilobites. *Geology* **27,** 987–990 (1999).
13. Chen, J.-Y., Ramsköld, L. & Zhou, G.-Q. Evidence for monophyly and arthropod affinity of Cambrian giant predators. *Science* **264,** 1304–1308 (1994).
14. Bergström, J. in *Darwin's Heritage Today. Proceedings of the Darwin 200 International Conference* 29–42 (Higher Education, 2010).
15. Hou, X.-G., Bergström, J. & Ahlberg, P. *Anomalocaris* and other large animals in the Lower Cambrian Chengjiang fauna of southwest China. *GFF* **117,** 163–183 (1995).
16. Usami, Y. Theoretical study on the body form and swimming pattern of *Anomalocaris* based on hydrodynamic simulation. *J. Theor. Biol.* **238,** 11–17 (2006).
17. Collins, D. The ''evolution'' of *Anomalocaris* and its classification in the arthropod class Dinocarida (nov.) and order Radiodonta (nov.). *J. Paleontol.* **70,** 280–293 (1996).
18. Daley, A. C., Budd, G. E., Caron, J.-B., Edgecombe, G. D. & Collins, D. The Burgess Shale anomalocaridid *Hurdia* and its significance for early euarthropod evolution. *Science* **323,** 1597–1600 (2009).
19. Kühl, G., Briggs, D. E. G. & Rust, J. A great-appendage arthropod with a radial mouth from the Lower Devonian Hunsrück Slate, Germany. *Science* **323,** 771–773 (2009).
20. Ma, X.-Y., Hou, X.-G. & Bergström, J. Morphology of *Luolishania longicruris* (Lower Cambrian, Chengjiang Lagerstätte, SW China) and the phylogenetic relationships within lobopodians. *Arthropod Struct. Dev.* **38,** 271–291 (2009).
21. Gehling, J. G., Jago, J. B., Paterson, J. R., García-Bellido, D. C. & Edgecombe, G. D. The geological context of the lower Cambrian (series 2) Emu Bay Shale Lagerstätte and adjacent stratigraphic units, Kangaroo Island, South Australia. *Aust. J. Earth Sci.* **58,** 243–257 (2011).
22. McKirdy, D. M. *et al.* Paleoredox status and thermal alteration of the lower Cambrian (Series 2) Emu Bay Shale Lagerstätte, South Australia. *Aust. J. Earth Sci.* **58,** 259–272 (2011).
23. Gabbott, S. E., Hou, X.-G., Norry, M. J. & Siveter, D. J. Preservation of Early Cambrian animals of the Chengjiang biota. *Geology* **32,** 901–904 (2004).
24. Gaines, R. R., Briggs, D. E. G. & Zhao, Y. Cambrian Burgess Shale-type deposits share a common mode of fossilization. *Geology* **36,** 755–758 (2008).
25. Nedin, C. The Emu Bay Shale, a Lower Cambrian fossil Lagerstätten, Kangaroo Island, South Australia. *Mem. Assoc. Austral. Palaeontol.* **18,** 31–40 (1995).
26. Cronin, T. W. & Porter, M. L. Exceptional variation on a common theme: the evolution of crustacean compound eyes. *Evol. Edu. Outreach* **1,** 463–475 (2008).
27. Wehner, R. in *Comparative Physiology and Evolution of Vision in Invertebrates: C* 287–616 (Springer, 1981).
28. Harzsch, S. Neurophylogeny: architecture of the nervous system and a fresh view on arthropod phylogeny. *Integr. Comp. Biol.* **46,** 162–194 (2006).
29. Vannier, J., García-Bellido, D. C., Hu, S.-X. & Chen, A.-L. Arthropod visual predators in the early pelagic ecosystem: evidence from the Burgess Shale and Chengjiang biotas. *Proc. R. Soc. Lond. B* **276,** 2567–2574 (2009).
30. Hu, S.-X. *et al.* Diverse pelagic predators from the Chengjiang Lagerstätte and the establishment of modern-style pelagic ecosystems in the early Cambrian. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **254,** 307–316 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. and C. Buck for access to and assistance at the fossil site; R. Atkinson, M. Gemmell, N. Schroeder, M. Binnie, J. Gehling, B. McHenry and others (Supplementary table 3 in ref. 1) for assistance with excavations and curation; A. Daley and J. Gehling for comments and advice; D. Birch and N. Vella for help with SEM-EDS analyses; and the Australian Research Council (grant LP0774959), South Australian Museum, University of Adelaide, MICINN-Spanish Ministry of Science (RYC2007-00090 and grant CGL2009-07073), Beach Energy and Sealink Pty Ltd for funding.

Author Contributions All authors directly contributed to excavation and interpretation of fossil specimens, analysis, and writing the manuscript. J.R.P. and D.C.G.-B. conducted the digital photography and camera lucida drawings; G.A.B. and J.R.P. conducted the SEM-EDS analyses.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.R.P. (jpater20@une.edu.au).

©2011 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature10571

Ecology drives a global network of gene exchange connecting the human microbiome

Chris S. Smillie[1]*, Mark B. Smith[2]*, Jonathan Friedman[1], Otto X. Cordero[3], Lawrence A. David[4] & Eric J. Alm[3,5,6]

Horizontal gene transfer (HGT), the acquisition of genetic material from non-parental lineages, is known to be important in bacterial evolution[1,2]. In particular, HGT provides rapid access to genetic innovations, allowing traits such as virulence[3], antibiotic resistance[4] and xenobiotic metabolism[5] to spread through the human microbiome. Recent anecdotal studies providing snapshots of active gene flow on the human body have highlighted the need to determine the frequency of such recent transfers and the forces that govern these events[4,5]. Here we report the discovery and characterization of a vast, human-associated network of gene exchange, large enough to directly compare the principal forces shaping HGT. We show that this network of 10,770 unique, recently transferred (more than 99% nucleotide identity) genes found in 2,235 full bacterial genomes, is shaped principally by ecology rather than geography or phylogeny, with most gene exchange occurring between isolates from ecologically similar, but geographically separated, environments. For example, we observe 25-fold more HGT between human-associated bacteria than among ecologically diverse non-human isolates ($P = 3.0 \times 10^{-270}$). We show that within the human microbiome this ecological architecture continues across multiple spatial scales, functional classes and ecological niches with transfer further enriched among bacteria that inhabit the same body site, have the same oxygen tolerance or have the same ability to cause disease. This structure offers a window into the molecular traits that define ecological niches, insight that we use to uncover sources of antibiotic resistance and identify genes associated with the pathology of meningitis and other diseases.

The human body is a complex biological network comprising ten microbes for each human cell and 100 microbial genes for each unique human gene[6]. Because this hidden microbial majority is known to have profound impacts on many aspects of human health including immunity[7], inflammatory disease[8] and obesity[9], considerable efforts are underway to document the genetic diversity of the human microbiome. The role of HGT in the generation and distribution of this biochemical repertoire is unclear, although anecdotal findings suggest that it may be significant[4,5,10]. In addition to informing our understanding of microbial evolution, predictive models of gene transfer are needed for the effective engineering of the human microbiome because HGT facilitates rapid adaptation to drugs and other perturbations[4,5]. Until now, however, a dearth of available genome sequences and appropriate analytical techniques have left an incomplete view of the forces that govern HGT[11].

Many previous efforts to explore these forces have highlighted the relationship between phylogeny and HGT[11–14]. Phylogeny is expected to influence HGT strongly because shared evolutionary history is associated with overlap in the host range of mobile elements[14], establishing a mechanistic basis for the phylogenetic control of gene exchange. Meanwhile, upon transfer, selection favours the persistence of genes acquired from close relatives, because these genes have greater compatibility with native molecular machinery[15,16].

Geography might provide an alternative structure to HGT by restricting dispersal, as suggested by the geographically organized distribution of *Vibrio cholera* integrons[17] and NDM-1 antibiotic resistance genes[18].

A third possibility is that ecological similarity shapes networks of gene exchange by selecting for the transfer and proliferation of adaptive traits or by increasing physical interactions between community members. Reports of enriched levels of HGT between hyperthermophiles[19] and spatially segregated exchange among *Shewanella* isolates[20] offer suggestive glimpses of such an ecological structure. However, it has been difficult to determine whether ecology has a broader function in HGT because of the limited availability of genomes from similar environments and because most previous work has ignored the distinction between recent transfers and ancient events. The inclusion of transfers from millions or billions of years in the past can obscure ecological structure, because historical niches may not reflect modern environmental associations.

To explore the effects of phylogeny, geography and ecology on HGT we use an evolutionary-rate heuristic to identify recent transfers between thousands of microbial genomes. Our heuristic finds blocks of nearly identical DNA (more than 500 nucleotides, more than 99% identity) in distantly related genomes (less than 97% 16S rRNA similarity). HGT is the best explanation for these observations because the highly conserved 16S gene evolves about 25-fold more slowly than protein-coding synonymous sites[21]. As a result, vertically inherited orthologues in such divergent genomes are nearly saturated with mutations at synonymous sites[22], in contrast to the almost perfect identity that we require. To avoid overcounting transfers, we cluster similar genomes and normalize against the number of possible comparisons.

We have confirmed that at least 98% of all HGT events identified with our approach include a predicted protein-coding gene, indicating that potentially problematic non-coding elements do not significantly affect our results. To validate our HGT detection method further, we use two phylogenetic inference methods to evaluate the evolutionary origins of putatively transferred sequences. Quartet mapping and a gene loss analysis each support 99% of identified HGTs (Supplementary Fig. 1).

As expected, a large fraction of observed transfers (27%) include at least one predicted mobile element, underscoring the importance of these genes in facilitating exchange. However, when we account for redundancies we find that mobile elements such as plasmids (2%), phages (1%) and transposons (9%) reflect only a promiscuous minority of the 10,770 total unique proteins that we observe, whereas the majority of unique genes (87%) provide other functions.

Direct exchange between any two bacteria in our data set is unlikely, both because we limit our analysis to distantly related bacteria and because strains were isolated from different human subjects or

[1]Computational and Systems Biology Initiative, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [2]Microbiology Graduate Program, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [3]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [4]Society of Fellows, Harvard University, Cambridge, Massachusetts 02138, USA. [5]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [6]Broad Institute, Cambridge, Massachusetts 02139, USA.
*These authors contributed equally to this work.

©2011 Macmillan Publishers Limited. All rights reserved

environments, often on different continents. An average pairwise distance of 7,000 km separates bacteria engaging in HGT. Therefore, each observed HGT probably reflects two independent acquisitions from a shared pool of mobile DNA, followed by proliferation.

To quantitatively explore the connectivity of bacteria in the human microbiome relative to other environments, we compare gene transfer between the 1,183 human-associated bacteria and 1,052 non-human-associated isolates from a broad range of aquatic, terrestrial and host-associated environments across the world. Even after correcting for biased sampling of human-associated clades (see Methods), pairs of bacteria isolated from the human body are 25-fold more likely to share transferred DNA than pairs from other environments ($P = 3.0 \times 10^{-270}$; combined Mann–Whitney *U*-test).

This enrichment in human-associated transfer may be caused by the prevalence of overlapping selective pressures in the tightly regulated, endothermic human host in comparison with diverse, non-human environments that experience significant temporal and spatial variation in selective pressures. Consistent with this hypothesis, when the environment is specified more precisely by focusing on human isolates from the same body site, we observe twofold higher rates of transfer ($P = 9.9 \times 10^{-108}$; combined Mann–Whitney *U*-test). Among the most closely related isolates from the same body site, this corresponds to recent HGT in more than 40% of comparisons. This elevated transfer between bacteria isolated from similar environments extends beyond the human body, with threefold more HGT between bacteria isolated from the same non-human environment relative to isolates from different non-human environments ($P = 1.3 \times 10^{-31}$; combined Mann–Whitney *U*-test).

However, an alternative explanation for these observations is that closely related bacteria colonize similar environments, creating an apparent ecological effect that is actually driven by shared evolutionary history. To control for such a phylogenetic effect, we plotted observed HGT over a range of phylogenetic divergences and found that the strong enrichment for exchange within similar environments (same host, same body site, same non-human environment) persisted across all distances (Fig. 1).

For a direct comparison of the relative contributions of phylogeny and ecology to the enrichment in human-associated transfer, we computed recent HGT between bacteria isolated from the human body (same ecology) and between these human-associated bacteria and all non-human-associated isolates (different ecology) over a range of phylogenetic distances. As shown by the dashed line in Fig. 2a, even the most deeply divergent bacteria that are separated by billions of years of evolution but share the same ecology engaged in more HGT than the mostly closely related isolates with different ecology. Thus, this recent gene exchange is structured by ecology more than by phylogeny.

We used a similar approach to explore the influence of geography relative to phylogeny and found that exchange between continents was slightly lower than exchange within the same continent (Fig. 2b; $P = 0.02$; combined Mann–Whitney *U*-test). However, this geographic effect was much weaker than that of phylogeny, which was itself less informative than ecology. Taken together, these analyses indicate that recent HGT frequently crosses continents and the Tree of Life to connect the human microbiome globally in an ecologically structured network.

This ecological architecture might reflect only the especially pronounced ecological differences between human-associated and non-human-associated bacteria. To determine whether ecology has a broad influence on recent gene exchange we searched for enriched HGT in narrower spatial, functional, and niche resolutions within the human host. Across all of these dimensions ecology strongly predicted gene exchange.

In addition to the previously discussed finding that transfer was enriched among bacteria from the same body site (Fig. 1), we found that further specifying the subsite of isolation (for example, separating vaginal isolates from other urogenital isolates) revealed even higher

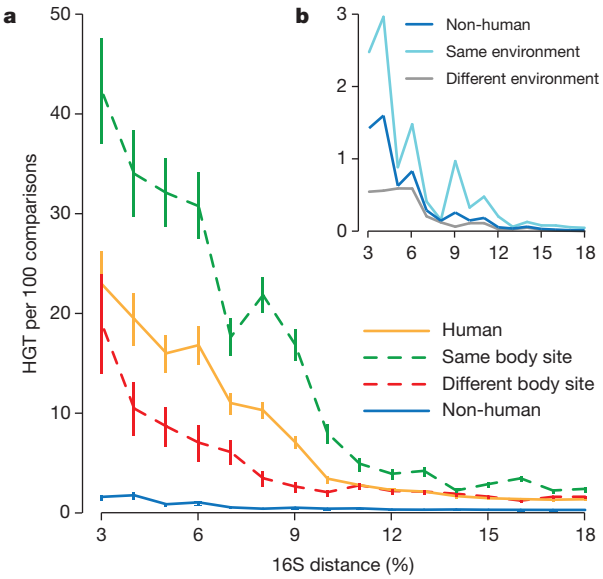


Figure 1 | Recent HGT is enriched in the human microbiome across all phylogenetic distances. HGT frequency is plotted as a function of the phylogenetic divergence between species for human-associated bacteria (**a**) and non-human-associated bacteria (**b**). We define species as clusters of genomes separated by less than 2% 16S rRNA divergence. HGT frequency is calculated in bins of 1% 16S rRNA divergence. Error bars indicate s.d. (see Supplementary Methods), with sample sizes described in Supplementary Table 8. These trends are also observed after controlling for the potential effects of sequencing centre contamination (Supplementary Fig. 4) and cosmopolitan strains (Supplementary Fig. 6).

levels of transfer across all three annotated body subsites (vagina, gingiva and nasopharynx) (Fig. 3a and Supplementary Figs 2 and 3; $P = 1.7 \times 10^{-9}$; combined Mann–Whitney *U*-test). When all human and non-human environments were considered, with scales ranging from tissues to ecosystems, we found that exchange at a narrow spatial scale, within an environment, always exceeded exchange at a broader spatial scale, with all other environments (Fig. 3b; $P = 1.3 \times 10^{-273}$; combined χ^2).

Up to this stage, our analysis relied on isolation environment as a proxy for ecological similarity, ignoring heterogeneities within these sites. Next we explored these differences, by evaluating the effects on HGT of oxygen tolerance and pathogenicity—the only other sufficiently annotated ecological features. Even after controlling for the effects of

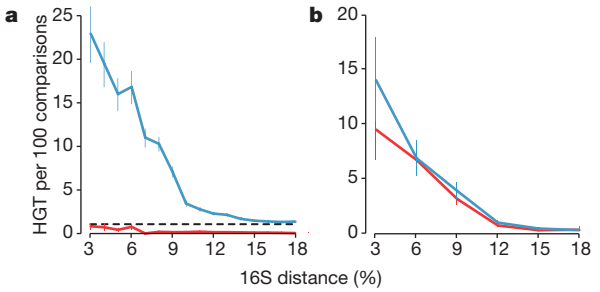


Figure 2 | Ecology is the dominant force shaping recent HGT in the human microbiome. **a**, The frequency of HGT between human-associated isolates (same ecology; blue) and between human-associated and non-human-associated isolates (different ecology; red). **b**, The frequency of HGT between bacteria isolated from the same continent (blue) and different continents (red). As a result of a reduced sample size in **b**, we pooled comparisons into larger phylogenetic distance bins of 3%. Error bars were calculated as in Fig. 1. The role of ecology in **a** is recovered when we control for sequencing centre contamination (see Supplementary Fig. 5).

©2011 Macmillan Publishers Limited. All rights reserved

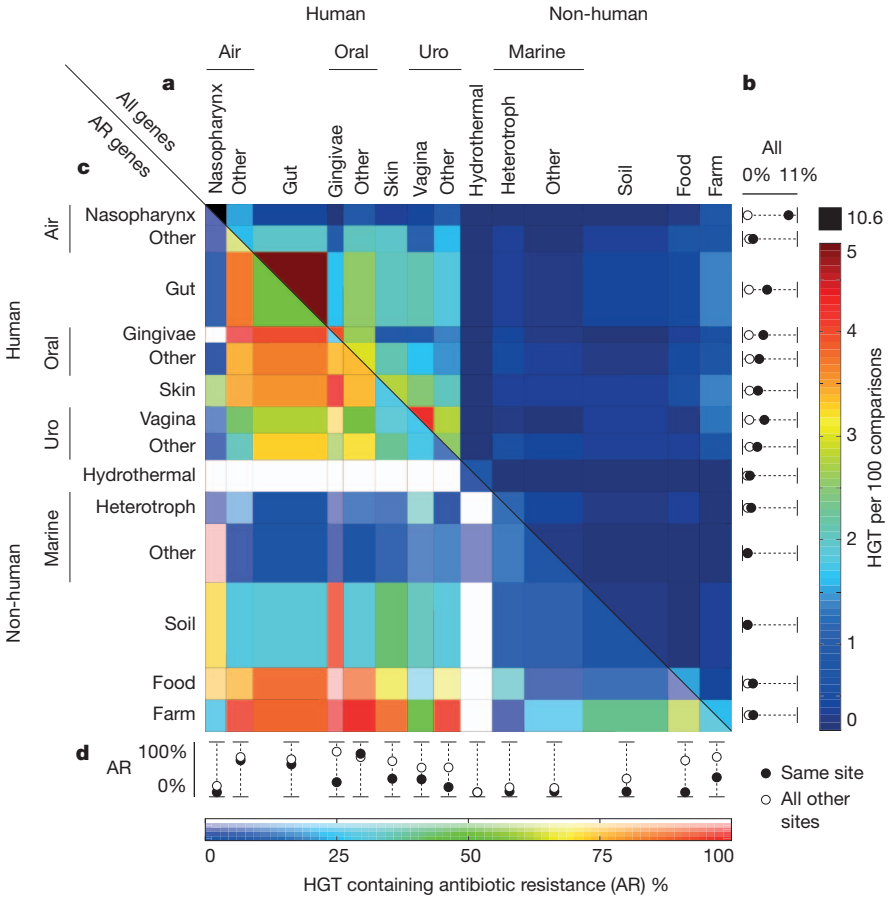


Figure 3 | HGT is ecologically structured by functional class and at multiple spatial scales. The frequency of transfer between different environments is shown for all functional groups (**a**, **b**) and for antibiotic resistance (AR) genes only (**c**, **d**). Box widths indicate the number of genomes from each environment. **a**, When all genes are considered (upper half), human isolates form a block of enrichment (upper left). **b**, For every environment examined we observe more transfer within the same environment (black dots) than between environments (white dots). **c**, The fraction of gene transfers that includes at least one AR gene for each environment. Statistical uncertainty in the proportion of AR transfer is indicated by decreased colour saturation (see Methods). **d**, AR genes comprise a significantly higher fraction of observed HGT between different environments (white dots) relative to that within the same environment (black dots) in contrast to **b**. Uro, urogenital.

body site and phylogeny, we found that HGT was also structured by oxygen tolerance (Fig. 4a; $P = 7.7 \times 10^{-13}$; χ^2) and pathogenicity (Fig. 4b; $P = 7.4 \times 10^{-11}$; χ^2). These findings demonstrate that in addition to the extensive spatial effects described above, chemical gradients and symbiotic relationships provide further ecological structure to recent HGT. Because these results persisted after controlling for explicit spatial effects, they seem to reflect selection rather than simply co-occurrence.

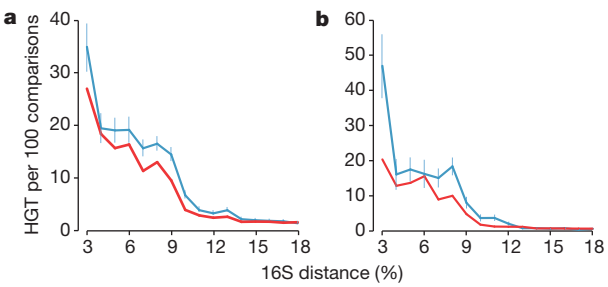


Figure 4 | Gene exchange is ecologically structured by oxygen tolerance and pathogenicity. The frequency of HGT between genomes with the same oxygen tolerance (**a**) and pathogenicity (**b**) is shown (blue) relative to their expected values (red). Expected values are based on overall frequencies of transfer between bacteria from the same distribution of body sites and phylogenetic distances. Bacteria that share the same oxygen tolerance (aerobic, anaerobic, microaerophilic or facultative aerobic) and pathogenicity (pathogenic or commensal) engage in significantly more HGT than expected under the null model, in which these traits have no influence on HGT. Error bars were calculated as in Fig. 1.

To further explore the role of selection, we probed its effects on the proliferation of different functional classes. If selection influences the rates and bounds of gene exchange, then the transfer of genes providing a non-specific selective advantage, such as antibiotic resistance, should show reduced environmental specificity relative to other, more niche-specific, functional classes. To test this prediction, for each environment, we considered the fraction of observed transfers that included at least one antibiotic resistance gene (Fig. 3c). In contrast to our earlier observation of increased transfer within sites when all functional classes were grouped together (Fig. 3a, b), here we observed that resistance comprised a higher fraction of transfers across different environments than within the same environment (Fig. 3d; $P = 6.9 \times 10^{-279}$; combined χ^2). Thus, when ecological forces transcend environmental boundaries, mobile genes do too.

We have explored networks of gene transfer to evaluate the forces that influence recent HGT, finding that ecology is profoundly important. Next we demonstrated how knowledge of this association between ecology and HGT could be used to reveal clinical insights from patterns of observed gene transfer.

Our findings, coupled with previous results[5], suggest that recently transferred genes between bacteria occupying a well-defined niche are especially likely to reflect adaptation to that niche. Consistent with this expectation, we found that many genes transferred between distantly

©2011 Macmillan Publishers Limited. All rights reserved

related meningitis isolates—such as hemolysins, adhesins and antibiotic resistance genes (Supplementary Table 1)—are known to be important in the disease[23]. We suggest that other transferred genes with unknown functions are probably cryptic virulence factors and should be prioritized for experimental annotation. Thus, in addition to recovering known virulence factors, our approach might streamline the search for novel drug targets[24], because although it is prohibitively difficult to explore all 24,095 unique meningitis genes with unknown function, it is tractable to evaluate the 13 that were recently transferred. We used this approach to identify genes associated with other diseases (for example pneumonia and endocarditis; Supplementary Tables 2 and 3) and environments (for example hot springs and soil; Supplementary Tables 4 and 5), opening a molecular window into the genetic traits that define ecological niches.

As a second example, our analysis of recent HGT revealed potential sources of clinical antibiotic resistance. We found that bacteria from farm animals and human food were enriched in transfer of resistance with human-associated bacteria relative to other non-human-associated isolates ($P = 1.7 \times 10^{-11}$ and $P = 0.01$, respectively; Mann–Whitney U-test). In all, 42 unique antibiotic resistance genes were transferred between human and farm isolates. These transferred genes comprised nine families, all of which included both genes known to provide resistance to clinical antibiotics and genes known to confer resistance to agricultural drugs (see Supplementary Table 6). This suggests that livestock-associated bacteria can contribute to clinical resistance without directly infecting humans, because for these mobile traits, genes, not genomes, serve as the unit of evolution and proliferation. Moreover, we observed 43 unique antibiotic resistance genes crossing national borders, suggesting that because the human microbiome is globally connected, local contamination of the shared mobile gene pool can have significant transnational consequences.

We have shown that ecology governs recent HGT and used this finding to reveal the key genes and networks of exchange that facilitate colonization, and occasionally exploitation, of the human host. In the future this approach could be extended to analyse bacterial genomes from individuals or groups of individuals that differ in diet, disease or descent to search for the microbial genes that relate to these human conditions.

METHODS SUMMARY

All 16S genes were identified with the GreenGenes database[25]. A total of 115 genomes with spurious or truncated 16S sequences were excluded from our analysis. We used BLAST (version 2.2.20) with default parameters[26] to calculate an all-against-all nucleotide alignment for 2,235 genomes downloaded from IMG[27]. We inferred HGT events from blocks of nearly identical DNA (more than 99% identity, more than 500 bp) in distantly related genomes (less than 97% 16S rRNA similarity). To avoid overcounting events in ancestral lineages, we collapsed closely related genomes by using average linkage clustering into groups ('species') with a 16S dissimilarity of 2%. For each pair of these clusters, we calculated the fraction of genome comparisons between clusters that shared at least one inferred HGT event. We summed this fraction over all pairs of clusters and normalized to the total number of comparisons, to calculate the HGT per 100 comparisons. Statistical tests of HGT enrichment were performed separately for each distance bin, then combined into a single p value with Fisher's method. We modelled antibiotic resistance transfer as a binomial random variable with parameter p and calculated a 95% confidence interval around our estimate of p. The size of this confidence interval, which is the statistical uncertainty of our estimate, was used to desaturate the colour of the heat map in Fig. 3c. To explore the effects of oxygen tolerance and pathogenicity on HGT, we used a χ^2 test to compare the observed frequency of HGT with the expected value given the distribution of body sites and phylogenetic divergences. Protein-coding regions were identified and annotated with BLASTX[26] (Expect (E) value $< 10^{-50}$) and UBLAST[28] (maxtargets = 100, E value $< 10^{-50}$) searches against the NCBI nr database. Unique genes reflect unique best BLAST hits to the database. Antibiotic resistance genes were annotated with the Antibiotic Resistance Genes Database[29]. Data sets are available from http://almlab.mit.edu/data.

Received 20 May; accepted 19 September 2011.
Published online 30 October 2011.

1. Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405,** 299–304 (2000).
2. Koonin, E. V., Makarova, K. S. & Aravind, L. Horizontal gene transfer in prokaryotes: quantification and classification. *Annu. Rev. Microbiol.* **55,** 709–742 (2011).
3. Chen, J. & Novick, R. P. Phage-mediated intergeneric transfer of toxin genes. *Science* **323,** 139–141 (2009).
4. Lester, C. H., Frimodt-Moller, N., Sorensen, T. L., Monnet, D. L. & Hammerum, A. M. *In vivo* transfer of the *vanA* resistance gene from an *Enterococcus faecium* isolate of animal origin to an *E. faecium* isolate of human origin in the intestines of human volunteers. *Antimicrob. Agents Chemother.* **50,** 596–599 (2006).
5. Hehemann, J.-H. *et al.* Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464,** 908–912 (2010).
6. Gill, S. R. *et al.* Metagenomic analysis of the human distal gut microbiome. *Science* **312,** 1355–1359 (2006).
7. Round, J. L. & Mazmanian, S. K. The gut microbiota shapes intestinal immune responses during health and disease. *Nature Rev. Immunol.* **9,** 313–323 (2009).
8. Xavier, R. J. & Podolsky, D. K. Unravelling the pathogenesis of inflammatory bowel disease. *Nature* **448,** 427–434 (2007).
9. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444,** 1022–1023 (2006).
10. Xu, J. *et al.* Evolution of symbiotic bacteria in the distal human intestine. *PLoS Biol.* **5,** e156 (2007).
11. Lawrence, J. G. & Hendrickson, H. Lateral gene transfer: when will adolescence end? *Mol. Microbiol.* **50,** 739–749 (2003).
12. Thomas, C. M. & Nielsen, K. M. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Rev. Microbiol.* **3,** 711–721 (2005).
13. Gogarten, J. P., Doolittle, W. F. & Lawrence, J. G. Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* **19,** 2226–2238 (2002).
14. Mazodier, P. & Davies, J. Gene transfer between distantly related bacteria. *Annu. Rev. Genet.* **25,** 147–171 (2011).
15. Tuller, T. *et al.* Association between translation efficiency and horizontal gene transfer within microbial communities. *Nucleic Acids Res.* **39,** 1–13 (2011).
16. Jain, R., Rivera, M. C. & Lake, J. A. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl Acad. Sci. USA* **96,** 3801–3806 (1999).
17. Boucher, Y. *et al.* Local mobile gene pools rapidly cross species boundaries to create endemicity within global *Vibrio cholerae* populations. *mBio* **2**(2), e00335–10, doi:10.1128/mBio.00335-10 (2011).
18. Kumarasamy, K. K. *et al.* Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. *Lancet Infect. Dis.* **10,** 597–602 (2010).
19. Aravind, L., Tatusov, R. L., Wolf, Y. I., Walker, D. R. & Koonin, E. V. Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends Genet.* **14,** 442–444 (1998).
20. Caro-Quintero, A. *et al.* Unprecedented levels of horizontal gene transfer among spatially co-occurring *Shewanella* bacteria from the Baltic Sea. *ISME J.* **5,** 131–140 (2010).
21. Ochman, H., Elwyn, S. & Moran, N. A. Calibrating bacterial evolution. *Proc. Natl Acad. Sci. USA* **96,** 12638–12643 (1999).
22. Ochman, H. & Wilson, A. C. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J. Mol. Evol.* **26,** 74–86 (1987).
23. Kim, K. S. Pathogenesis of bacterial meningitis: from bacteraemia to neuronal injury. *Nature Rev. Neurosci.* **4,** 376–385 (2003).
24. Clatworthy, A. E., Pierson, E. & Hung, D. T. Targeting virulence: a new paradigm for antimicrobial therapy. *Nature Chem. Biol.* **3,** 541–548 (2007).
25. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72,** 5069–5072 (2006).
26. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
27. Markowitz, V. M. *et al.* The integrated microbial genomes (IMG) system. *Nucleic Acids Res.* **34,** D344–D348 (2006).
28. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,** 2460–2461 (2010).
29. Liu, B. & Pop, M. ARDB—Antibiotic Resistance Genes Database. *Nucleic Acids Res.* **37,** D443–D447 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by National Science Foundation awards 0918333 and 0936234 to E.J.A., and by the Department of Energy's ENIGMA Scientific Focus Area. This work is part of the National Institutes of Health Human Microbiome Project.

Author Contributions C.S.S., M.B.S. and E.J.A. conceived the study. C.S.S., M.B.S., J.F. and E.J.A. analysed the data. C.S.S., M.B.S., J.F., O.X.C., L.A.D. and E.J.A. provided conceptual insight. C.S.S., M.B.S. and E.J.A. prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to E.J.A. (ejalm@mit.edu).

©2011 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature10555

Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome

Claude Becker[1]*, Jörg Hagmann[1]*, Jonas Müller[1], Daniel Koenig[1], Oliver Stegle[2], Karsten Borgwardt[2] & Detlef Weigel[1]

Heritable epigenetic polymorphisms, such as differential cytosine methylation, can underlie phenotypic variation[1,2]. Moreover, wild strains of the plant *Arabidopsis thaliana* differ in many epialleles[3,4], and these can influence the expression of nearby genes[1,2]. However, to understand their role in evolution[5], it is imperative to ascertain the emergence rate and stability of epialleles, including those that are not due to structural variation. We have compared genome-wide DNA methylation among 10 *A. thaliana* lines, derived 30 generations ago from a common ancestor[6]. Epimutations at individual positions were easily detected, and close to 30,000 cytosines in each strain were differentially methylated. In contrast, larger regions of contiguous methylation were much more stable, and the frequency of changes was in the same low range as that of DNA mutations[7]. Like individual positions, the same regions were often affected by differential methylation in independent lines, with evidence for recurrent cycles of forward and reverse mutations. Transposable elements and short interfering RNAs have been causally linked to DNA methylation[8]. In agreement, differentially methylated sites were farther from transposable elements and showed less association with short interfering RNA expression than invariant positions. The biased distribution and frequent reversion of epimutations have important implications for the potential contribution of sequence-independent epialleles to plant evolution.

Although there is no doubt that DNA sequence mutations are the primary raw material for evolutionary change, local DNA methylation variants with major effects on the expression of nearby genes can be inherited over many generations[1,2]. However, such epialleles are not always as stable as the primary DNA sequence[3,9–11]. New sequencing technologies have recently enabled the direct determination of spontaneous DNA mutation rates[12], and we have previously reported that *A. thaliana* experiences about one single-base-pair mutation per haploid genome and generation[7]. This analysis was based on a set of five mutation accumulation lines that had been derived from a single individual of the inbred strain used to produce the high-quality reference genome sequence for *A. thaliana*. These lines had been separately propagated in a common environment by single-seed descent for 30 generations[6]. We examined whole-genome cytosine methylation[13,14] in these five lines plus five additional lines of the same population by Illumina sequencing. We interrogated two siblings each of the 31st generation with an average strand-specific coverage depth of 20× per individual; changes shared within a line should predominantly reflect differences that had accumulated by the 30th generation. Because seeds from the founders were no longer available, we compared the 31st generation individuals to two independent lines that had been propagated for only three generations (Supplementary Fig. 1).

Out of all cytosine residues with high-quality sequencing support (see Supplementary Methods), on average 2.8 million were found to be methylated in each line (Supplementary Table 1). The higher genome-wide methylation rate in our analysis compared to previous studies[13,14] reflects the greater statistical power afforded by increased sequencing depth. We subsequently evaluated 13.9 million cytosines that had at least threefold coverage in all individuals, of which 3 million were methylated in at least one strain. Using Fisher's exact test, we identified about 186,000 (6.2%) positions with a significant change in methylation (false discovery rate <0.05) between at least one 31st generation and both 3rd generation lines. Almost all, 99.6%, of these differentially methylated positions (DMPs) were also detected with an entropy-based method[15]. Given the limited statistical power for weakly methylated or poorly covered sites (Supplementary Figs 2–4), our DMP estimate would almost certainly increase with higher sequencing depth. For further analyses, we considered sites that agreed between 31st generation siblings (on average, 99.8%) and between the two strains closest to the founder generation (99.7%).

CG sites were highly over-represented among DMPs (Fig. 1a). This is unlikely to reflect greater instability of CG compared to CHG and CHH positions (where H is A, T or C), but rather higher statistical power in detecting a change at CG sites, which are on average much more highly methylated[13,14] (Supplementary Fig. 4). Among CG sites in genic regions, including those producing non-coding RNAs, relative abundance of DMPs was two- to fourfold higher compared with non-differentially methylated positions (N-DMPs). The opposite was the case for CG positions in transposable elements and intergenic regions, with a similar, but less pronounced, bias for CHG and CHH sites (Fig. 1b). These observations were in agreement with CG-DMPs being found most often on chromosome arms, which have the highest gene density (Fig. 1c), even though cytosine methylation near the centromeres is the highest[13,14]. Gene body methylation gradually increases towards the 3′ end, before sharply decreasing at the end of the last exon[13,14,16,17], although genes 1 kb or less in length were generally only weakly methylated (Supplementary Fig. 5a). The profiles of DMPs and N-DMPs were similar across individual genes, exons, introns and transposable elements (Fig. 1d and Supplementary Fig. 5b, c), but DMPs were less frequent in promoter and downstream regions. Notably, CG-DMPs accounted for 42% of methylated sites in gene bodies, despite all CG-N-DMPs outnumbering CG-DMPs four to one.

Twenty-four-nucleotide-long small interfering RNAs (siRNAs) are important in maintaining DNA methylation[8], and N-DMPs coincided seven times more often than DMPs with sites to which 24-nt siRNAs mapped[18]. N-DMPs were also on average only half as far from such sites as DMPs ($P < 2.2 \times 10^{-16}$) (Fig. 1e and Supplementary Fig. 6a). siRNAs are enriched in and around transposable elements[19]. In agreement, the average distance to the closest transposable element was much shorter for N-DMPs outside of transposable elements, compared to DMPs ($P < 2.2 \times 10^{-16}$), even when only considering those in the centromere-distant regions of each chromosome, which contain relatively few transposable elements (Fig. 1e and Supplementary Fig. 6b, c).

A first major insight from our analyses is that transgenerational maintenance of CG methylation in transposable elements is apparently

[1]Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany. [2]Machine Learning and Computational Biology Research Group, Max Planck Institute for Developmental Biology and Max Planck Institute for Intelligent Systems, 72076 Tübingen, Germany.
*These authors contributed equally to this work.

©2011 Macmillan Publishers Limited. All rights reserved

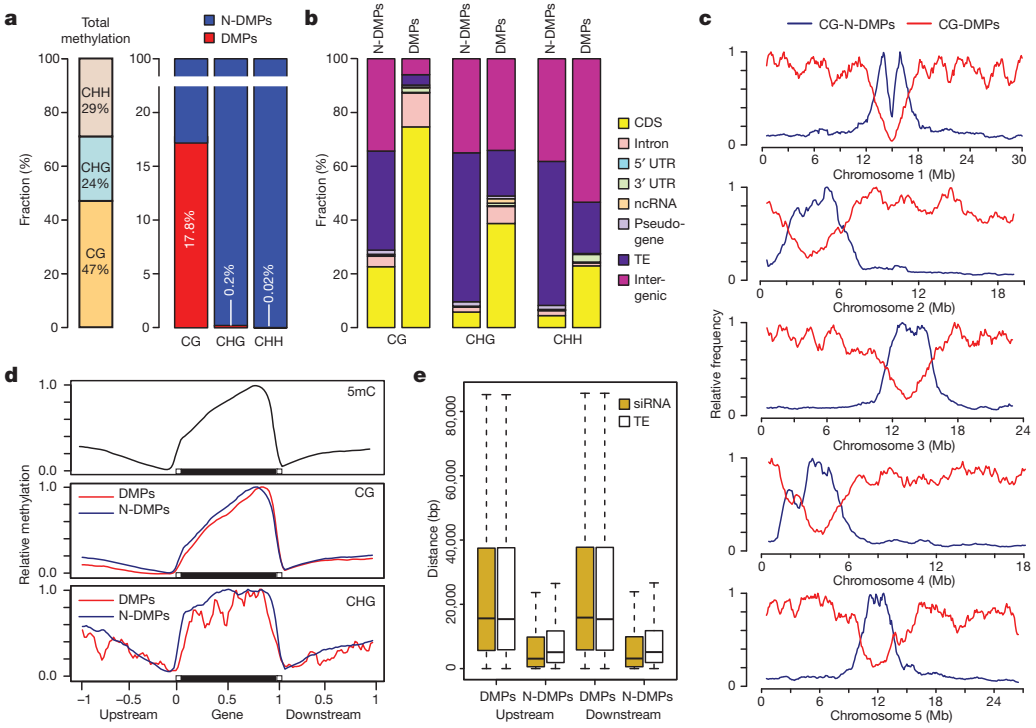


Figure 1 | Genome-wide distribution of methylation polymorphisms. **a**, Contribution of CG, CHH and CHG sites to total and differential cytosine methylation. 32.8% of all CG, 15.7% of CHG and 4.6% of CHH sites, adding up to 10.8% of all cytosines, showed evidence of methylation. **b**, Distribution of DMPs and N-DMPs according to local annotation. CDS, coding sequence; ncRNA, non-coding RNA; TE, transposable element. **c**, Distribution of CG-N-DMPs and CG-DMPs along each chromosome. Data were normalized to the highest value for each chromosome and class. **d**, Averaged distribution of all methylated sites (5mC) and methylated CG and CHG sites along genes. Data were normalized to the highest value for each sequence context and class. The coding region is indicated by a black bar. **e**, Distance of DMPs and N-DMPs to the closest upstream and downstream 24-nucleotide siRNA and transposable element. Horizontal bar corresponds to median, whiskers indicate entire 75th percentile.

much more stable than CG methylation of protein-coding genes, consistent with DNA methylation being more important for controlling the activity of transposable element compared to the latter[14,17,20,21]. This also agrees with a report that genic methylation is much more variable between wild strains of *A. thaliana* than methylation of transposable elements[3].

Hierarchical clustering based on DMPs grouped siblings as well as 3rd and 31st generation lines together. An arbitrary selection of methylated positions, which included about 6% DMPs, produced a similar pattern; however, with N-DMPs only, clusters became much more random (Fig. 2a). These observations indicate that our DMPs capture most of the methylation differences between lines. We next calculated the pairwise distance between strains based on DMPs (Fig. 2b). Correlation was highest between the two 3rd generation strains, and each individual of the 31st generation was more similar to these two lines, from which they were separated by 34 generations, than to the other lines from the 31st generation, from which they had diverged for 62 generations. Taken together, we conclude that whole-genome methylation patterns are largely stable and therefore heritable in *A. thaliana*, but that differences in methylation status accumulate gradually, similar to genetic mutations.

One strain, 69, was exceptional and had 40% more DMPs in comparison with the 3rd generation than the other 31st generation lines (Fig. 2b). To determine whether this strain might have a defect in the methylation machinery, we sequenced its genome with more and longer reads compared to our previous analysis[7]. We found a non-synonymous change in *MATERNAL EFFECT EMBRYO ARREST 57* (*MEE57*), which encodes a protein related to METHYLTRANSFERASE 1 (MET1) (Supplementary Fig. 7). *MEE57* has been reported as essential for endosperm development[22], although several *A. thaliana* strains lack functional *MEE57* copies[23]. Thus, whether the *MEE57* mutation contributes to the increased DMP number in line 69 remains unclear. The fact that the siblings of this line were as similar to each other as other sibling pairs (Fig. 2a) argues against a generally increased epimutation rate.

Compared to genetic mutations, the frequency of epimutations at single cytosine residues was many orders of magnitude higher, with an average of close to 30,000 DMPs in the analysed sequence space, compared with less than 30 DNA sequence mutations per strain[7]. Thirty-two per cent of DMPs between generations 3 and 31 occurred more than once, and 13% more than twice (Fig. 2c). If DMPs arose randomly, we would expect less than 1% of recurrent events. That we observe many more indicates that certain positions are particularly prone to increases or decreases in methylation rate. To investigate directly how many DMPs emerge from one generation to the next, we analysed the 32nd generation of lines 39 and 49. These individuals were progeny of siblings of the individuals interrogated in the 31st generation, and shared changes in the 32nd generation should reflect differences that arose between the 30th and 31st generation. We found on average over 3,300 between-generation DMPs. This is in the same range as DMPs between siblings (on average, about 5,000), but more than we would have expected from the 30,000 that had accumulated between the 3rd and each of the 31st generation lines. One explanation is that frequent transgenerational changes in methylation status occur at a limited number of sites, and that only a fraction of new DMPs is maintained over the longer term. This is corroborated by the observation that more than two-thirds of DMPs distinguishing the 32nd from the 31st generation in lines 39 and 49 had already been found in other 31st generation individuals.

DNA methylation is known to occur nonrandomly, and to cluster in specific segments of the genome[3,17,24]. We identified 249 differentially methylated regions (DMRs) that were at least 50 bp long (median 100 bp, maximum 650 bp) (Supplementary Table 2 and Supplementary Methods). Although probably a conservative estimate, the number of DMRs per line is in the same range as the DNA sequence mutations,

©2011 Macmillan Publishers Limited. All rights reserved

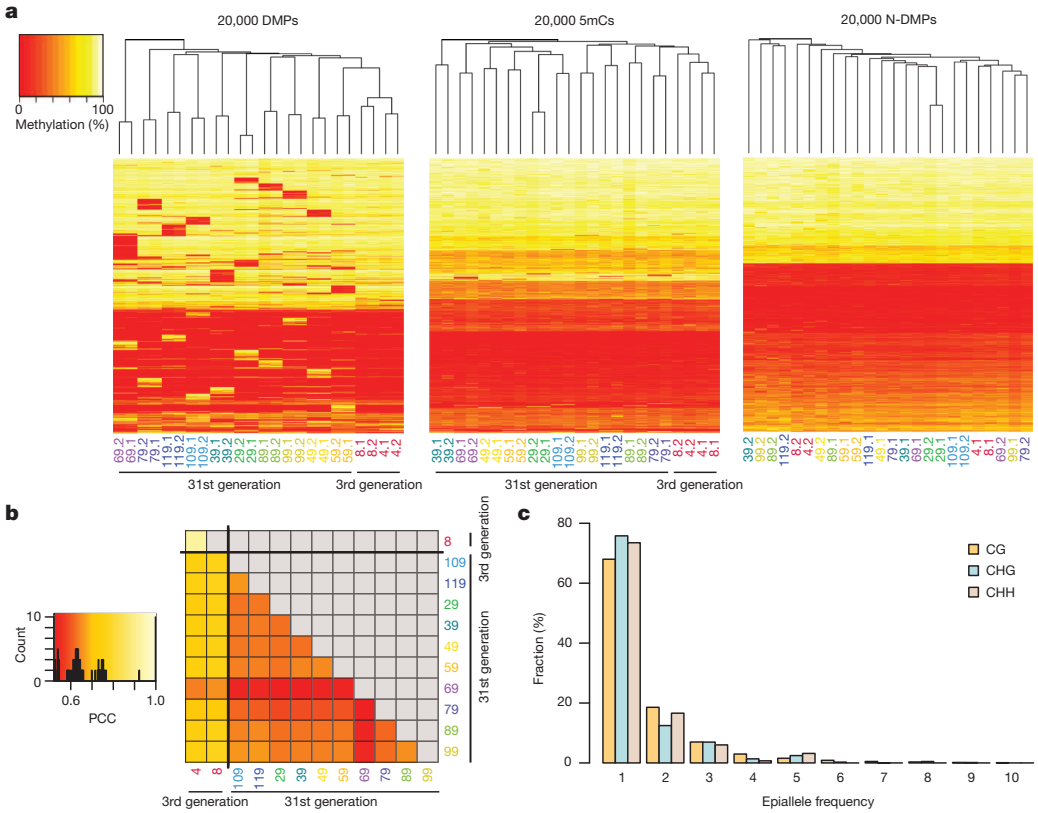


Figure 2 | Epigenetic diversity in the analysed population. **a**, Hierarchical clustering based on 20,000 sites each, drawn randomly from DMPs identified in pairwise comparison between all strains, cytosines methylated in at least one of the analysed strains (including about 6% DMPs), or N-DMPs. **b**, Heat map representing pairwise Pearson's correlation coefficient (PCC) between individuals, considering all 250,000 DMPs identified between all strains. PCCs between 3rd generation strains, 0.92; between 3rd and 31st generation, 0.63–0.77; between 31st generation lines, 0.52–0.66. The histogram on top of the colour key indicates counts of PCC bins. **c**, Epiallele frequency of DMPs in the 31st generation.

less than 30 per line[7]. As with CG-DMPs, DMRs preferentially localized to genes (Fig. 3a). DMRs did not overlap with known DNA mutations in these strains[7]. Similarly, structural variant discovery with established methods[23,25] did not reveal evidence for DMRs being due to gross DNA lesions. The frequency of DMRs along genes was similar to the overall distribution of methylated cytosines, and was reminiscent of the pattern of variation seen in wild strains of *A. thaliana*[3] (Supplementary Fig. 8). There were almost ten times as many DMRs in exons as in introns (Fig. 3a). Because exon-specific methylation may influence RNA splicing patterns[26,27], this could also be a source of variation in gene activity. Hierarchical clustering according to DMRs separated early- and late-generation strains into distinct groups (Fig. 3b). Notably, if we consider the methylation status in the 3rd generation individuals as largely reflecting the ancestral pattern, similar fractions of DMRs had lost or gained methylation by the 31st generation (Fig. 3c).

Similarly to DMPs, recurrent events constituted more than one-third of all DMRs, indicating that the affected genomic regions were privileged sites of change (Fig. 3d, e). In addition, comparison of generations 32 and 31 identified four short DMRs per line, with re-methylation of one segment that had become unmethylated in generation 31 (Supplementary Fig. 9). Together, these observations demonstrate that large changes in methylation, although rare, can occur even within a single generation.

Differences in promoter and genic DNA methylation can affect RNA levels[16]. We compared the transcriptomes of two randomly selected strains of the 31st generation with the 3rd generation strains by RNA-seq (Supplementary Fig. 10) and identified 320 differentially expressed genes in pairwise comparisons between strains (Fig. 3f and Supplementary Table 3). The two 31st generation lines were separated from each other by the most changes, and the two 3rd generation lines by the fewest. Seven differentially expressed genes overlapped with a DMR in these strains (Fig. 3g). For the three genes with the highest difference in expression level and overlapping with the most conspicuous DMRs, we observed a negative correlation between DNA methylation and gene expression. The remaining four genes overlapped with much shorter DMRs and no correlation was apparent (Fig. 3e, g and Supplementary Fig. 11).

We have presented a high-resolution analysis of transgenerational variation in DNA methylation of *A. thaliana*. The molecular mechanisms underlying these changes remain elusive, but siRNAs, which map very often in or near transposable elements[19], probably have a role in stabilizing DNA methylation, which is corroborated by our finding that DMPs tend to be farther from transposable elements and to be associated with lower local siRNA activity than N-DMPs. These observations indicate that the density and distribution of transposable elements, which can differ greatly even between closely related species[28], affect epigenetic variation throughout the genome. In the material analysed here, there was no evidence for DNA mutations acting *in cis* as an important cause of DMRs, although we cannot rule out that a non-synonymous mutation in a *MET1* homologue might contribute to increased variation in DNA methylation *in trans* in one of the lines.

In contrast to the high frequency of single-nucleotide methylation polymorphisms, larger regions appear to change methylation status at a rate that is comparable to genetic mutations. On the basis of previous work[7], it is conceivable that the emergence of DMRs requires specific structural features such as nearby repeats. Although DMRs are rare, we found evidence for DMRs affecting gene expression, indicating that natural, sequence-independent epialleles could potentially contribute to phenotypic diversity. There are subtle morphological differences between the mutation accumulation lines[6], and quantitative genetic approaches could be used to link specific DNA mutations or DMRs with such traits. How many of the methylation differences found

©2011 Macmillan Publishers Limited. All rights reserved

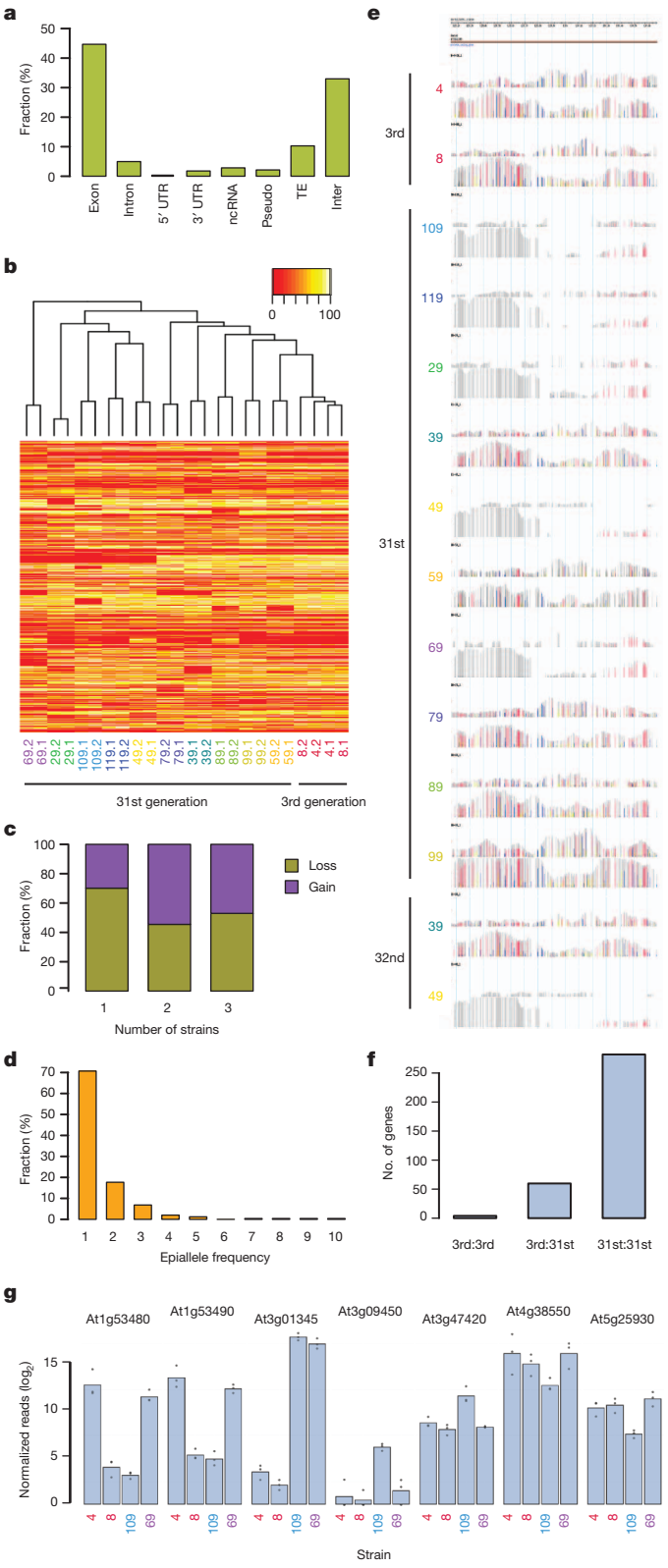


Figure 3 | Differentially methylated regions (DMRs). **a**, Distribution of DMRs according to local annotation. Inter, intergenic; pseudo, pseudogene. **b**, Hierarchical clustering of individuals from the 3rd and 31st generation based on methylated sites in DMRs, ranked according to their position in the genome. Note the shared methylation differences across lines, but strict pairing of siblings. **c**, Regions with losses and gains of methylation in 1, 2 or 3 strains in generation 31 compared to the 3rd generation strains. **d**, Epiallele frequency of DMRs. **e**, DMR at At3g01345 (Chr3:129,159–130,670) across all strains. Methylation on both strands is indicated for each strain. Colours indicate methylated reads (red, CG; blue, CHG; yellow, CHH). Grey indicates reads supporting non-methylation. For simplicity, only one sibling is shown per strain. **f**, Differentially expressed genes in comparisons between 3rd and 31st generation strains. **g**, Average RNA expression levels of the genes overlapping with the regions in **e** and in Supplementary Fig. 11. Dots indicate values of individual samples.

between wild strains[3] are due to sequence-independent changes versus ones driven by transposable elements and other structural variants is an important area for further investigation. In addition, it will be necessary to follow DNA methylation not only under benign greenhouse conditions, but also in the much more variable and stressful natural environment.

Perhaps our most important finding is that the number of epimutations does not increase linearly with time, indicating that many are not stably inherited over the long term. In addition to DMPs and DMRs that arose apparently independently in several strains, we even discovered a DMR that had become demethylated after 31 generations, but was re-methylated in the following generation. This suggests that DNA methylation in specific regions of the genome can fluctuate over relatively short timescales. Such sites can be considered as going through recurrent cycles of forward and reverse epimutation, which is very different from what is found at the level of the genome sequence, where reverse mutations are exceedingly rare. Importantly, reversion rates directly determine the ability of any type of allele to be subject to Darwinian selection. This needs to be taken into account when considering the potential of epialleles as a factor in evolution[5].

METHODS SUMMARY

Methylome sequencing. DNA was prepared from nuclei isolated from leaf tissue, bisulphite treated using a modification of a published protocol[14], and paired-end sequenced on the Illumina GAIIx platform. After image analysis and base calling with the Illumina pipeline, reads were processed using SHORE[7,29], and aligned to the Col-0 reference genome with GenomeMapper[30], adapted to the analysis of bisulphite sequencing data. Bisulphite conversion rates, as determined from unmethylated chloroplast and spiked-in lambda phage DNA, were 99.72% to 99.84%.

Analysis of methylated positions. Single sites were classified as methylated or unmethylated by fitting a binomial model based on reads falsely reporting methylation on the unmethylated plastid genome. We only considered cytosine residues covered by at least three independent high-quality base calls in all strains. For the determination of significant differences in methylation across strains on single sites or regions, we used Fisher's exact test.

Data visualization. A Gbrowse instance of the methylation profiles is available at http://gbrowse.weigelworld.org/fgb2/gbrowse/ath_methyl_ma.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 July; accepted 13 September 2011.
Published online 20 September 2011.

1. Richards, E. J. Population epigenetics. *Curr. Opin. Genet. Dev.* **18,** 221–226 (2008).
2. Paszkowski, J. & Grossniklaus, U. Selected aspects of transgenerational epigenetic inheritance and resetting in plants. *Curr. Opin. Plant Biol.* **14,** 195–203 (2011).
3. Vaughn, M. W. *et al.* Epigenetic natural variation in *Arabidopsis thaliana*. *PLoS Biol.* **5,** e174 (2007).
4. Rangwala, S. H. *et al.* Meiotically stable natural epialleles of Sadhu, a novel *Arabidopsis* retroposon. *PLoS Genet.* **2,** e36 (2006).
5. Slatkin, M. Epigenetic inheritance and the missing heritability problem. *Genetics* **182,** 845–850 (2009).
6. Shaw, R. G., Byers, D. L. & Darmo, E. Spontaneous mutational effects on reproductive traits of *Arabidopsis thaliana*. *Genetics* **155,** 369–378 (2000).
7. Ossowski, S. *et al.* The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* **327,** 92–94 (2010).
8. Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Rev. Genet.* **11,** 204–220 (2010).
9. Teixeira, F. K. *et al.* A role for RNAi in the selective correction of DNA methylation defects. *Science* **323,** 1600–1604 (2009).
10. Reinders, J. *et al.* Compromised stability of DNA methylation and transposon immobilization in mosaic *Arabidopsis* epigenomes. *Genes Dev.* **23,** 939–950 (2009).
11. Widman, N., Jacobsen, S. E. & Pellegrini, M. Determining the conservation of DNA methylation in *Arabidopsis*. *Epigenetics* **4,** 119–124 (2009).
12. Lynch, M. Evolution of the mutation rate. *Trends Genet.* **26,** 345–352 (2010).

©2011 Macmillan Publishers Limited. All rights reserved

13. Cokus, S. J. *et al.* Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452,** 215–219 (2008).
14. Lister, R. *et al.* Highly integrated single-base resolution maps of the epigenome in *Arabidopsis. Cell* **133,** 523–536 (2008).
15. Zhang, Y. *et al.* QDMR: a quantitative method for identification of differentially methylated regions by entropy. *Nucleic Acids Res.* **39,** e58 (2011).
16. Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T. & Henikoff, S. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genet.* **39,** 61–69 (2007).
17. Zhang, X. *et al.* Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis. Cell* **126,** 1189–1201 (2006).
18. Fahlgren, N. *et al.* MicroRNA gene evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana. Plant Cell* **22,** 1074–1089 (2010).
19. Kasschau, K. D. *et al.* Genome-wide profiling and analysis of *Arabidopsis* siRNAs. *PLoS Biol.* **5,** e57 (2007).
20. Slotkin, R. K. *et al.* Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136,** 461–472 (2009).
21. Lisch, D. Epigenetic regulation of transposable elements in plants. *Annu. Rev. Plant Biol.* **60,** 43–66 (2009).
22. Pagnussat, G. C. *et al.* Genetic and molecular identification of genes required for female gametophyte development and function in *Arabidopsis. Development* **132,** 603–614 (2005).
23. Cao, J. *et al.* Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nature Genet.* doi:10.1038/ng.911 (28 August 2011).
24. Tran, R. K. *et al.* DNA methylation profiling identifies CG methylation clusters in *Arabidopsis* genes. *Curr. Biol.* **15,** 154–159 (2005).
25. Mirouze, M. *et al.* Selective epigenetic control of retrotransposition in *Arabidopsis. Nature* **461,** 427–430 (2009).
26. Chodavarapu, R. K. *et al.* Relationship between nucleosome positioning and DNA methylation. *Nature* **466,** 388–392 (2010).
27. Laurent, L. *et al.* Dynamic changes in the human methylome during differentiation. *Genome Res.* **20,** 320–331 (2010).
28. Hollister, J. D. *et al.* Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata. Proc. Natl Acad. Sci. USA* **108,** 2322–2327 (2011).
29. Ossowski, S. *et al.* Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res.* **18,** 2024–2033 (2008).
30. Schneeberger, K. *et al.* Simultaneous alignment of short reads against multiple genomes. *Genome Biol.* **10,** R98 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to R. Shaw for providing seeds of the mutation accumulation lines. We thank C. Lanz for help with Illumina sequencing, J. Fitz for help with the Gbrowse implementation, K. Schneeberger for help with DNA mutation analysis, D. Zilberman for providing bisulphite sequencing protocols, and R. Clark, M. Pellegrini, R. Neher and J. Paszkowski for discussion. This work was supported by an HFSP Long-Term Fellowship (D.K.), the Volkswagen Foundation (O.S.), FP7 Collaborative Project AENEAS (contract KBBE-2009-226477), a Gottfried Wilhelm Leibniz Award of the DFG, and the Max Planck Society (D.W.).

Author Contributions C.B., J.H. and D.W. conceived the study; C.B. performed the experiments; C.B., J.H., J.M., D.K. and O.S. analysed the data; K.B. provided advice on statistical analysis; and C.B. and D.W. wrote the paper with contributions from all authors.

Author Information DNA sequencing data are deposited in the European Nucleotide Archive (http://www.ebi.ac.uk/ena/data/view/ERP000902) under accession number ERP000902. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.W. (weigel@weigelworld.org).

©2011 Macmillan Publishers Limited. All rights reserved

METHODS

Plant growth and material. Seeds were derived from *Arabidopsis thaliana* Columbia-0 lines in generation 3 (lines 4 and 8), generation 31 (lines 29, 39, 49, 59, 69, 79, 89, 99, 109 and 119) and generation 32 (lines 39 and 49), counting from the founders, as described by Shaw and colleagues[6] (Supplementary Fig. 1). Plants were grown on soil under long-day conditions (23 °C, 16 h light, 8 h dark) after seeds had been stratified in 150 nM GA-supplemented water at 4 °C for 6 days. Siblings were grown independently at different time points. Positions of the pots were randomized.

Nucleic acid extraction. DNA was extracted from rosettes of individual 21-day-old plants. Plant material was flash-frozen in liquid nitrogen and ground in a mortar. The ground tissue was re-suspended in nuclei extraction buffer (10 mM Tris-HCl pH 9.5, 10 mM EDTA, 100 mM KCl, 0.5 M sucrose, 0.1 mM spermine, 0.4 mM spermidine, 0.1% β-mercaptoethanol). After cell lysis in nuclei extraction buffer with 10% Triton X-100, nuclei were pelleted by centrifugation at 2,000*g* for 120 s. Genomic DNA was extracted using the Qiagen Plant DNeasy kit (Qiagen). Total RNA was extracted from rosette leaves of individual plants using the Trizol (Invitrogen) method according to the manufacturer's instructions. Residual DNA was eliminated by DNase I (Thermo Fisher Scientific) treatment.

Library preparation. Preparation of DNA libraries for genomic sequencing was done using the NEBNext DNA Sample Prep Reagent Set 1 (New England Biolabs), following the Illumina Genomic Sample Prep Guide (Illumina). 500–1,000 ng genomic DNA was fragmented to 300 bp average size with a Covaris S2 instrument using the following settings for 120 s in frequency sweeping mode: intensity 5, duty cycle 10%, 200 cycles per burst. DNA was purified on Qiaquick PCR purification columns. Preparation of DNA libraries for bisulphite sequencing was adapted from ref. 14. Input DNA was fragmented as described above. Libraries were constructed using the NEBNext DNA Sample Prep Reagent Set 1 (New England Biolabs) according to the Illumina Genomic Sample Prep Guide with the following modifications. We used the Illumina Early Access Methylation Adapter Oligo Mix (catalogue number ME-100-0010). After size selection, the non-methylated cytosine residues were converted to uracil using the EpiTect Plus DNA Bisulfite kit (Qiagen) according to the manufacturer's guidelines. For higher conversion efficiency the bisulphite incubation was repeated. Library enrichment was performed with Pfu Cx HotStart Polymerase (Agilent) and 18 PCR cycles. Libraries for RNA sequencing were prepared from 4 μg of total RNA using the Illumina Truseq RNA sample prep kit B according to the manufacturer's protocol.

Sequencing. All sequencing was performed on an Illumina GAIIx instrument. Genomic and bisulphite-converted libraries were sequenced with 2×101-bp paired-end reads. For bisulphite sequencing, conventional *A. thaliana* DNA genomic libraries were analysed in control lanes. Transcriptome libraries were sequenced with 101-bp single end reads, with three libraries with different indexing adapters pooled in one lane; no control lane was used. For image analysis and base calling, we used the Illumina OLB software version 1.8.

Processing and alignment of bisulphite-treated reads. The SHORE pipeline[29] was used to trim and quality-filter the reads. Its default parameters were applied for the filtering step: reads with more than 2 (or 5) bases in the first 12 (or 25) positions with a quality score less than 3 were discarded. Reads were trimmed to the right-most occurrence of two adjacent bases with quality values equal to or greater than 5. Trimmed reads shorter than 50 bases were discarded. The remaining high quality sequences (on average 82% of raw reads across the sequenced strains) were aligned against the *Arabidopsis thaliana* genome sequence version TAIR9 (http://www.arabidopsis.org/portals/genAnnotation/gene_structural_annotation/annotation_data.jsp) using a modified version of the mapping tool GenomeMapper[30] that supports the alignment of bisulphite converted reads. Bisulphite converts non-methylated cytosines into uracils, which are propagated as adenine-thymine base pairs after PCR amplification. GenomeMapper tolerates asymmetrical T-to-C or A-to-G mismatches (read base against reference base) and can distinguish between reads from the bisulphite-converted strand of a DNA fragment and sequences from its complementary amplified strand, if the reads have been obtained by paired-end sequencing. Only the read from the strand with converted Cs is informative about the methylation status of the underlying cytosine site. We allowed for up to 10% single-base-pair substitutions along the read length in the alignment process for each read to retain most coverage. GenomeMapper reports all alignments with the least amount of mismatches for each read. However, only reads mapping uniquely to a single position were used for this study. Furthermore, all but one read were removed from further analysis if their 5′ ends aligned to the same genomic position, to account for amplification biases. A paired-end correction method[23] was used to discard repetitive reads by comparing the distance between reads and their partner to the average distance between all read pairs. Reads with abnormal distances were removed if there was at least one other alignment of this read in a concordant distance to its partner. Finally, read counts on all cytosine sites were obtained with SHORE. The 'scoring matrix approach' of SHORE[23] assigns a score to each site by testing against different sequence and alignment related features. The criteria and complete scoring matrix can be found in Supplementary Table 4. For comparisons across lines, cytosines were accepted if at most one intermediate penalty on its score was applicable to at least one strain (score ≥32). In this case, the threshold for the other strains was lowered, accepting at most one high penalty (score ≥ 15). In this way, information from other strains is used to assess sites from the focal strain under the assumption of mostly conserved methylation patterns, allowing the analysis of additional sites. The methylation statistics on each single strain assumed a quality score of 25 or higher, which means no more than two intermediate penalties.

Determination of methylated sites. Sequencing errors, noise and imperfections of the bisulphite conversion contribute to the occurrence of sites that appear to be weakly methylated. Reads mapping against the non-methylated chloroplast sequence allow for objective estimation of the effective background rate of false-positive methylation detection. For this purpose, we fitted an independent binomial model to the relative proportions of converted and unconverted reads that cover cytosines in the chloroplasts. We estimated the binomial rate of false-positive methylation from the maximum likelihood estimate, separately for each library and for different bins of total read coverage:

$$\hat{r} = \arg\max_r \prod_s \text{Binomial}(\overline{n}_s, (n_s + \overline{n}_s) | r).$$

Here, n_s and $\overline{n}_s$ denote the number of converted and unconverted reads from the considered cytosine sites. Supplementary Fig. 12a shows the obtained background methylation rate for a single strain, line 30-39, as a function of the total read coverage per site. The overall false methylation rate when combining read data across the range of read coverage was 0.22%, which deviates significantly from higher error estimates when considering low-coverage regions in isolation. To account for the variability in error rates in the downstream analyses, we used specific error models for each strain and for read-coverage bins of multiples of fivefold, yielding error rates between 0.2% and 5.0% (Supplementary Fig. 12b). For coverage bins with too few sites for robust statistical estimation (<50), we imputed the false methylation rate from the closest sufficiently populated coverage bin. Given the estimated rates for false methylation, we carried out a genome-wide test for significant methylation of cytosines. For each site, we calculated the *P* value under the background model. We then used Storey's method[31], an extension of the Benjamini–Hochberg stepdown procedure, to assess genome-wide significance using *q* values. At a joint false discovery rate (FDR) of 5% we found between 2,316,966 and 3,458,949 methylated sites in each strain (Supplementary Table 1). When reducing FDR to 0.1%, we still retained almost 85% of the methylated sites, showing that the number of sites with weak methylation evidence was low. For analysis of methylated sites reported in this study, an FDR of 5% was deemed to be acceptable.

Identification of differentially methylated positions. From the 13.9 million cytosines for which we had at least three independent high-quality reads in each strain, we selected sites that showed significant methylation in at least one strain, resulting in 3,067,017 positions. Sites with statistically significant methylation differences were identified with Fisher's exact test. *P* values from individual tests per site were combined into single *P* values via conservative Bonferroni correction. Genome-wide FDRs were then estimated using Storey's method[31]. To limit false-positive DMPs, we first identified 69,583 DMPs between siblings at a relaxed FDR of 10%. These sites were excluded as were 8,893 DMPs distinguishing the two 3rd generation strains. This left 2,988,541 positions as the final set to test for differential methylation between generations. Twenty pairwise tests of each of the ten 31st generation strains against both 3rd generation strains were conducted on sites consistently methylated between 31st generation siblings and in the 3rd generation. At an FDR of 5%, this yielded 186,248 DMPs. DMP allele frequency was obtained by progressively removing the strain with the lowest *q* value and correcting the remaining *P* values for multiple testing by the methods described above.

We applied the same strategy to identify DMPs that differed either between the 31st and 3rd generation, or between 31st generation strains. Count data from replicates were combined for each site, followed by pairwise Fisher's exact tests between all combinations of strains (66 tests). We estimated *P* values for at least one differential pair using a Bonferroni correction, followed by Storey's method[31] to assess genome-wide significance. At a joint FDR of 5%, this identified 253,546 DMPs across all 12 strains.

Assessing statistical power. Two main factors influence the power to detect methylation differences: the number of statistical tests and local read coverage. To assess the impact of multiple testing, we applied the approach described above to all sites with at least threefold coverage in at least 12 of 24 individuals examined. Of 25.3 million such positions, 4,547,568 were found to be methylated in at least one of the lines, compared to 3,067,017 out of 13.9 million positions when considering only sites with complete information. The number of sites assessed as methylated thus increased roughly linearly with the number of tested sites, as did

©2011 Macmillan Publishers Limited. All rights reserved

the number of differentially methylated positions. Similarly, the fraction of DMPs shared in more than one 31st generation strain, ~31%, was very similar to the ~32% found among sites with complete information. We conclude that our method is largely insensitive to the number of tests performed.

To assess the effect of read coverage, we determined how many DMPs could be identified after subsampling at 25%, 50% and 75% of total coverage. We identified almost twice as many DMPs with 50% compared to 25% coverage, but only 13% additional DMPs were identified when increasing coverage from 75% to 100% (Supplementary Fig. 13). Although not yet asymptotic, we estimate that the false-negative rate is well below 50%, and most likely closer to 10%.

Identification of differentially methylated regions. The ~186,000 DMPs distinguishing 31st from both 3rd generation lines were consolidated into regions of adjacent DMPs for each strain, with a maximum distance of 50 bp between DMPs. We then used Fisher's exact test on the sum of methylated and unmethylated reads in both siblings, averaged across positions within the region. Resulting *P* values were corrected with Storey's method[31] and an FDR of 5% was accepted. Statistically significant regions from different strains were merged if they overlapped by at least 20% of their combined length and if the methylation change was in the same direction compared to the 3rd generation lines. Short regions containing only a small number of strongly differential sites were excluded by requiring DMRs to have a minimum length of 50 bp and to contain at least ten methylated positions and at least five DMPs.

Mapping of DMPs, N-DMPs and DMRs to genomic elements. We used the TAIR10 annotation (http://www.arabidopsis.org/portals/genAnnotation/gene_structural_annotation/annotation_data.jsp) to determine overlap of genes, pseudogenes and transposable elements with methylated positions. We defined intergenic regions as regions that did not correspond to any annotated feature. A DMR was considered mapping to a particular genomic element if it overlapped with such an element for more than 20% of its length.

Assessing the distance to the closest siRNA or transposable element. We determined the distance between a methylated position and the closest upstream and downstream siRNA using a published data set for *A. thaliana* aerial tissue[18] (NCBI GEO accession number GSM518432). For transposable elements, we used the TAIR10 annotation. Statistical significance was tested with a two-sided, unpaired Student's *t*-test on the measured distances. Pericentromeric regions were defined as described[23].

Analysis of gene expression. DNA sequences resulting from RNA library preparation were barcode-sorted and quality filtered in SHORE[29], and aligned using BWA[32] to the TAIR10 gene annotation. Reads were filtered for duplicates and required to have a mapping quality of at least 37. The remaining mappings were used to generate gene-level counts for expression analysis. We only considered genes for which the total counts in all samples combined exceeded 30. Between-sample expression correlations and strain-distribution plots were used for quality control to identify poor samples, and pairwise comparisons of expression were performed using the DEseq package[33] implemented in R. Differentially expressed genes were identified by a combination of per-gene variance ($P \leq 0.01$, with Benjamini and Yekutieli correction[34]) and common variance ($\geq 2\times$ change). Density of genes without expression support were plotted along chromosomes in sliding windows considering only genes with at least 50 methylated or unmethylated calls. Windows of gene methylation were calculated for entire gene bodies as the fraction of methylated positions (methylated in any sample) divided by the total number of called positions. Very similar results were obtained considering sites methylated in all samples. Data were visualized with the ggplots package in R[35].

Data visualization. A Gbrowse instance of the methylation profiles is available at http://gbrowse.weigelworld.org/fgb2/gbrowse/ath_methyl_ma.

31. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100,** 9440–9445 (2003).
32. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
33. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).
34. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29,** 1165–1188 (2001).
35. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009).

©2011 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature10665

Predicting mutation outcome from early stochastic variation in genetic interaction partners

Alejandro Burga[1], M. Olivia Casanueva[1] & Ben Lehner[1,2]

Many mutations, including those that cause disease, only have a detrimental effect in a subset of individuals. The reasons for this are usually unknown, but may include additional genetic variation and environmental risk factors[1]. However, phenotypic discordance remains even in the absence of genetic variation, for example between monozygotic twins[2], and incomplete penetrance of mutations is frequent in isogenic model organisms in homogeneous environments[3,4]. Here we propose a model for incomplete penetrance based on genetic interaction networks[5,6]. Using *Caenorhabditis elegans* as a model system, we identify two compensation mechanisms that vary among individuals and influence mutation outcome. First, feedback induction of an ancestral gene duplicate differs across individuals, with high expression masking the effects of a mutation. This supports the hypothesis that redundancy is maintained in genomes to buffer stochastic developmental failure[7]. Second, during normal embryonic development we find that there is substantial variation in the induction of molecular chaperones such as Hsp90 (DAF-21). Chaperones act as promiscuous buffers of genetic variation[8,9], and embryos with stronger induction of Hsp90 are less likely to be affected by an inherited mutation. Simultaneously quantifying the variation in these two independent responses allows the phenotypic outcome of a mutation to be more accurately predicted in individuals. Our model and methodology provide a framework for dissecting the causes of incomplete penetrance. Further, the results establish that inter-individual variation in both specific and more general buffering systems combine to determine the outcome inherited mutations in each individual.

It is well established that isogenic individuals show substantial variation at the molecular level, for example in the expression levels of particular genes[10]. This variation can be an important influence on signalling and development in wild-type (WT) individuals[11,12]. Such molecular variation may also influence the phenotypic consequences of inherited mutations, but examples of molecular variation have only been described that predict phenotypic variation at the level of other molecules[13] or cellular phenotypes[14].

To identify determinants of mutation outcome, we present a model for incomplete penetrance based on genetic interaction networks[5,6] (Fig. 1b). Our model proposes that in the absence of additional genetic variation, it is stochastic variation in the abundance or activity of genetic interaction partners (genes that influence the outcome of a mutation when genetically altered) that determines the outcome of a mutation (Fig. 1d). The correlated loss of genetic, environmental and stochastic robustness upon gene deletion in yeast is highly consistent with the generality of this proposal[6,15].

To test our model, we first used a null mutation in the T-box transcription factor gene *tbx-9* (Supplementary Fig. 1) that causes an incompletely penetrant defect in *C. elegans* larval morphology owing to abnormal development of the epidermis and muscle[16,17] (Fig. 1a). Transcription factor *tbx-9* is related to another transcription factor, *tbx-8*, by an ancestral gene duplication; inactivation of *tbx-8* causes similar incompletely penetrant defects, and loss of both genes results in synthetic lethality in *C. elegans*[16,17] and *Caenorhabditis briggsae*[18].

If the incomplete penetrance of *tbx-9* relates to variation in the expression of *tbx-8*, then increased expression of *txb-8* should be able to compensate for the loss of *tbx-9*. Expression of TBX-8 from a transgene indeed reduces the penetrance of a *tbx-9*(*ok2473*) null allele (Fig. 1c and Supplementary Fig. 2). A further requisite of our model is the existence of inter-individual variation in the expression of the genetic interaction partner. We quantified the induction of a fluorescent reporter driven by the *tbx-8* promoter (*ptbx-8::GFP*) during normal development and found that there was substantial variation among embryos (Fig. 2a) with an important extrinsic component (Supplementary Figs 3a and 4).

We next quantified the induction of the *tbx-8* reporter in *tbx-9*(*ok2473*) mutant animals and found that expression was increased (Fig. 2b, 1.2-fold upregulation at comma stage, $P = 1.6 \times 10^{-3}$). We confirmed this upregulation by performing single molecule fluorescence *in situ* hybridization against the endogenous *tbx-8* messenger RNA (mRNA) (Supplementary Fig. 5, 1.6-fold upregulation in 40- to

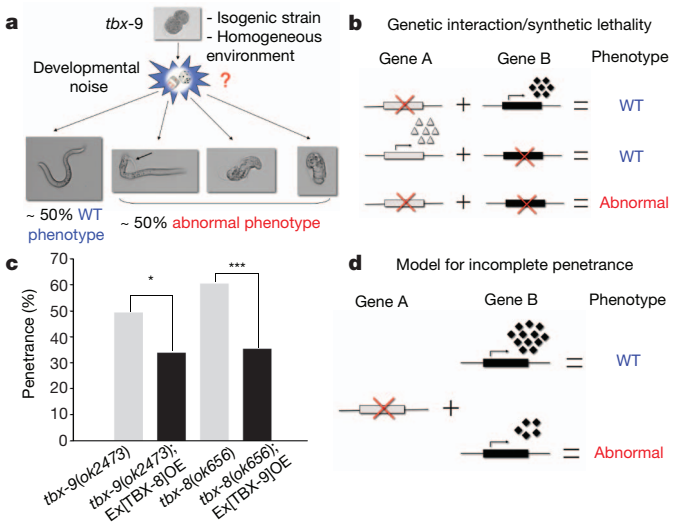


Figure 1 | Genetic interactions provide a general model for incomplete penetrance. **a**, Inactivation of the gene *tbx-9* in *C. elegans* results in an incompletely penetrant defect, with approximately half of embryos hatching with abnormal morphology (small arrow). **b**, Representation of a negative (synergistic) genetic interaction between two genes *A* and *B*. **c**, Increased expression of TBX-8 reduces the penetrance of a *tbx-9*(*ok2473*) null mutation (49% ($n = 138$) compared with 34% ($n = 139$), a 33% decrease in abnormal phenotypes, $P = 0.011$, Fisher's exact test). Likewise, increased TBX-9 expression rescues the penetrance of a *tbx-8*(*ok656*) null mutation (60% ($n = 164$) compared with 35% ($n = 195$), a 42% decrease, $P = 2.8 \times 10^{-6}$). **d**, A model for incomplete penetrance based on variation in the activity of genetic interaction partners.

[1]EMBL-CRG Systems Biology Unit, Centre for Genomic Regulation (CRG) and Universitat Pompeu Fabra (UPF), Barcelona 08003, Spain. [2]Institució Catalana de Recerca i Estudis Avançats, Centre for Genomic Regulation (CRG) and Universitat Pompeu Fabra (UPF), Barcelona 08003, Spain.

©2011 Macmillan Publishers Limited. All rights reserved

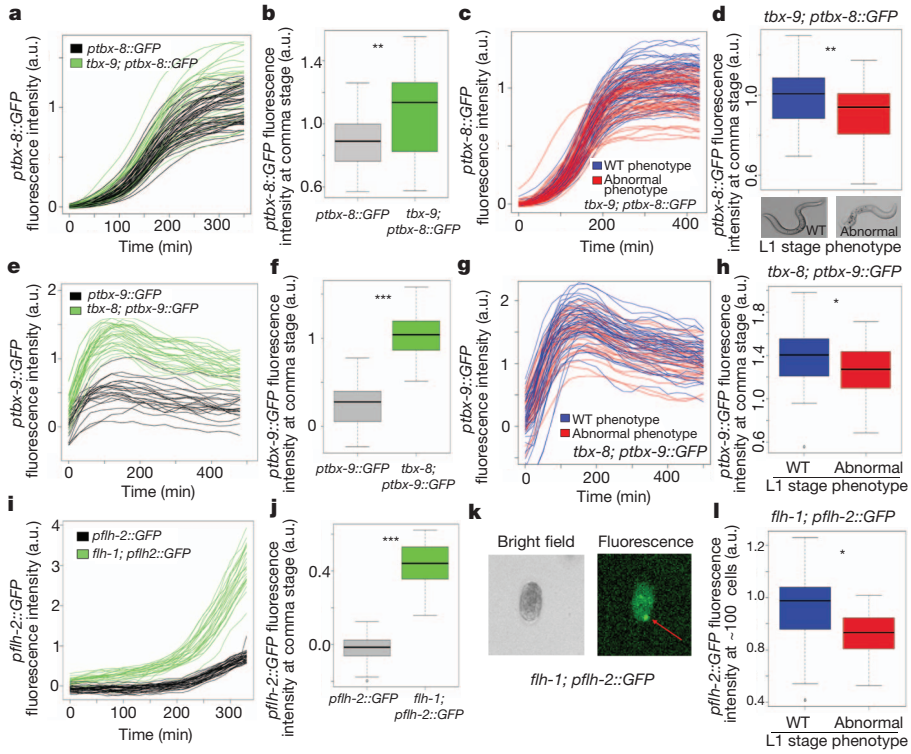


Figure 2 | Early inter-individual variation in the induction of ancestral gene duplicates predicts the outcome of inherited mutations. **a**, Quantification of total green fluorescent protein (GFP) expression from a *tbx-8* reporter during embryonic development in WT (black) and *tbx-9*(*ok2473*) (green) individuals. Each individual is a separate line. a.u., Arbitrary units. **b**, Boxplot of *tbx-8* reporter expression (**a**) showing 1.2-fold upregulation in a *tbx-9* mutant at comma stage (~290 min, $P = 1.6 \times 10^{-3}$, Wilcoxon rank test). **c**, Expression of *tbx-8* reporter in a *tbx-9*(*ok2473*) background for embryos that hatch with (red) or without (blue, WT) a morphological defect. **d**, Boxplot of **c** showing *tbx-8* expression is higher in *tbx-9* embryos that develop a WT phenotype (blue) compared with those that develop an abnormal (red) phenotype at comma stage ($P = 6.1 \times 10^{-3}$). **e**, Expression of a *ptbx-9::GFP* reporter in WT (black) and *tbx-8*(*ok656*) mutant (green). **f**, Boxplot of *tbx-9* reporter showing 4.3-fold upregulation at comma stage (~375 min, $P = 3.6 \times 10^{-16}$). **g**, Expression of *tbx-9* reporter in a *tbx-8*(*ok656*) mutant background, colour code as in **c**. **h**, Boxplot of **g** showing *tbx-9* expression is higher in *tbx-8* embryos that develop a WT phenotype ($P = 0.033$). **i**, Expression of a *pflh-2::GFP* reporter in WT (black) and *flh-1*(*bc374*) mutant (green). **j**, Boxplot of *flh-2* reporter expression (**i**) showing 1.8-fold upregulation in a *flh-1* mutant at comma stage (~180 min, $P = 2.2 \times 10^{-16}$). **k**, Bright-field and fluorescence image of an approximate 100-cell *flh-1*; *pflh-2::GFP* embryo. Red arrow indicates the local expression of *flh-2* reporter quantified for *flh-1* phenotypic prediction. **l**, Boxplot showing higher *flh-2* reporter expression at approximate 100 cells for WT (blue) compared with abnormal (red) phenotypes ($P = 0.014$). Boxplots show the median, quartiles, maximum and minimum expression in each data set.

50-cell-stage embryos, $P = 5.3 \times 10^{-4}$). Thus, a direct or indirect feedback mechanism exists that upregulates transcription of *tbx-8* when its ancestral paralogue is inactivated. This feedback also acts on the *tbx-9* promoter, which is also upregulated in a *tbx-9*(*ok2473*) mutant background (Supplementary Fig. 6). This is highly consistent with observations that have been made in yeast[19,20] and suggests that compensatory expression by negative feedback regulation of gene duplicates is probably a conserved phenomenon across species.

Next, we tested whether variation in the induction of *tbx-8* correlated with the outcome of the *tbx-9*(*ok2473*) mutation at hatching. We retrospectively compared the early expression of the reporter in embryos that did, and did not, hatch with an abnormal phenotype. Early expression was higher in the second class (Fig. 2c, d, $P = 6.1 \times 10^{-3}$); whereas 80% of the embryos with reporter expression in the highest quartile hatched without a defect, only 40% of those in the lowest quartile showed no phenotype (Supplementary Table 1, $P = 3.3 \times 10^{-3}$). In contrast, early expression from a ubiquitously transcribed promoter (*plet-858::GFP*) did not predict phenotypic outcome in individuals (Supplementary Fig. 7a, b, $P = 0.58$), nor did variation in the expression of a reporter for the transcription factor ELT-5, another protein required for epidermal development[21] (Supplementary Fig. 7c, d, $P = 0.97$). High expression of the synthetic lethal partner *tbx-8*, but not high transcription in general, is therefore partly epistatic[22] to the loss of *tbx-9*. We also tested the reverse situation: that is, whether variation in *tbx-9* expression compensation (Fig. 2e, f, 4.3-fold upregulation, $P = 3.6 \times 10^{-16}$) predicted the phenotypic outcome of a *tbx-8*(*ok656*) mutation. We found that was indeed the case (Fig. 2g, h, $P = 0.033$, see Supplementary Table 1).

The FLYWCH transcription factors *flh-1* and *flh-2* are an additional pair of ancestral gene duplicates that redundantly repress embryonic expression of the microRNAs *lin-4*, *mir-48* and *mir-241* (ref. 23). Similar to the case for *tbx-9* null mutants, *flh-1*(*bc374*) mutant embryos induced higher levels of a *pflh-2::GFP* reporter than WT embryos (Fig. 2i, j, 1.8-fold at comma stage, $P < 2.2 \times 10^{-16}$). In contrast, we found neither upregulation of a *tbx-8* reporter in an *flh-1*(*bc374*) background, nor upregulation of an *flh-2* reporter in a *tbx-9*(*ok2473*) mutant (Supplementary Fig. 8, $P = 0.74$ and $P = 0.55$ respectively). *flh-1*(*bc374*) embryos that expressed higher *flh-2* reporter levels early in development (Fig. 2k, see Methods) were, however, more likely to develop into morphologically WT larvae (Fig. 2l, $P = 0.014$, see Supplementary Table 1). In contrast, we found that *pflh-2::GFP* reporter levels did not predict the outcome of a *tbx-9*(*ok2473*) null mutation (Supplementary Fig. 7g, h, $P = 0.32$), nor did induction of the *ptbx-8::GFP* reporter predict the outcome of the *flh-1*(*bc374*) mutation (Supplementary Fig. 7e, f, $P = 0.49$).

Many ancestral gene duplicates have retained partly redundant functions over extensive evolutionary periods[19,24,25]. One explanation for this could be the selection pressure provided by stochastic developmental errors[7,26] ('canalization'[27]). Our results provide direct empirical support for this hypothesis, showing that when one member of a partly

©2011 Macmillan Publishers Limited. All rights reserved

redundant gene pair is inactivated, expression variation in the other gene becomes an important influence on the phenotype of an individual.

The variation in the induction *tbx-8* that we can quantify only partly accounts for variation in the outcome of the *tbx-9* mutation (Fig. 2d). We considered that one additional influence on mutation outcome could be variation in the activity of general buffering systems such as molecular chaperones. In multiple eukaryotes, chaperone inhibition can enhance the effects of diverse mutations[8,9,28] and increased chaperone activity promiscuously suppresses detrimental mutations in bacteria[29,30].

We constructed a transcriptional reporter for the constitutively expressed chaperone *daf-21* (*Hsp90*). Variability in the *pdaf-21::mCherry* reporter was evident during all larval stages and, surprisingly, its induction during embryonic development also varied substantially, even in the absence of a mutation or environmental perturbation (Supplementary Figs 9a and 10). No difference in transgene copy number was detected between animals with high or low *daf-21* reporter expression (Supplementary Fig. 11), and individuals with high reporter expression have high expression of the DAF-21 protein (Fig. 3a and Supplementary Fig. 12).

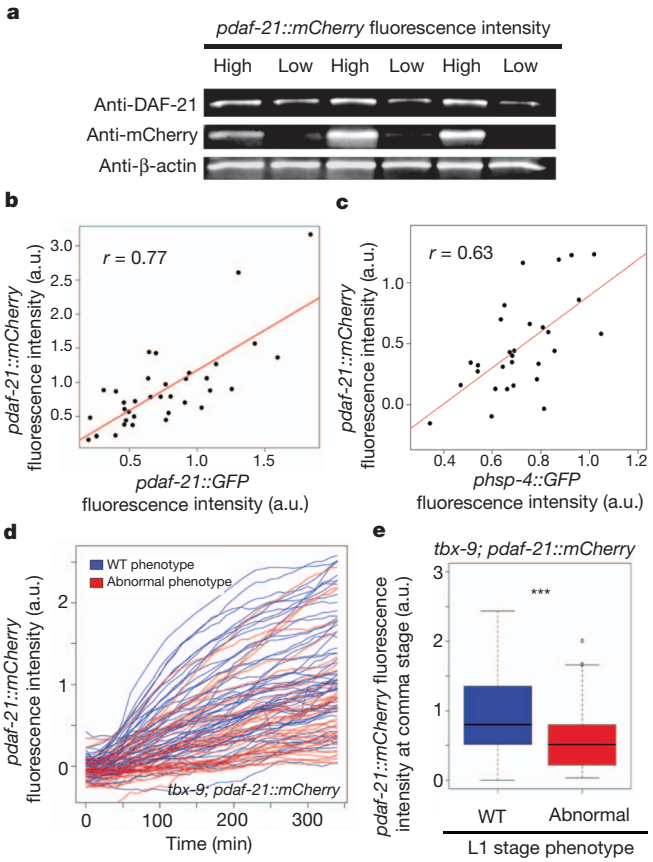


Figure 3 | Inter-individual variation in chaperone induction predicts the outcome of a mutation. **a**, *pdaf-21::mCherry; glp-1(e2141)* animals were sorted into 'high' and 'low' groups based on their *pdaf-21::mCherry* expression 24 h after the L1 stage at 25 °C. Western blot analysis confirmed that worms belonging to the 'high' group had higher levels of endogenous DAF-21 protein. As a positive control, mCherry protein levels showed the same trend. β-Actin was used as loading control; results of three replica sortings are shown. **b**, **c**, Correlation in individuals (measured at comma stage) between: *pdaf-21::mCherry* and *pdaf-21::GFP* (Pearson correlation coefficient, $r = 0.77$, $P = 3.4 \times 10^{-8}$) (**b**); *pdaf-21::mCherry* and *phsp-4::GFP* ($r = 0.63$, $P = 2.8 \times 10^{-4}$) (**c**). **d**, **e**, Expression of *daf-21* reporter in a *tbx-9(ok2473)* mutant background. Embryos that hatch into phenotypically WT worms (blue) have higher expression than those hatching with a morphological defect (red) at the comma stage ($P = 1.9 \times 10^{-3}$).

We tested whether this variability was particular to *daf-21* or common to other chaperones using a reporter for the gene *hsp-4* (orthologous to mammalian BiP). Transcription from the *hsp-4* promoter also varied extensively among embryos (Supplementary Fig. 13) and was correlated with that from the *daf-21* promoter (Fig. 3c, Pearson correlation coefficient, $r = 0.63$, $P = 2.8 \times 10^{-4}$), similar to the correlation between two independent *daf-21* reporters (Fig. 3b, $r = 0.77$, $P = 3.4 \times 10^{-4}$). This and additional correlations between stress response genes (Supplementary Fig. 14) suggest that, during normal development, isogenic embryos differ in a coordinated manner in the transcriptional induction of multiple chaperones.

We quantified induction of a *pdaf-21::mCherry* reporter in a *tbx-9(ok2473)* background. The mean induction of this reporter was increased in *tbx-9(ok2473)* animals (Supplementary Fig. 9b, 1.8-fold upregulation, $P = 6.6 \times 10^{-4}$). In addition, embryos that later hatched without a morphological defect expressed higher levels of the reporter

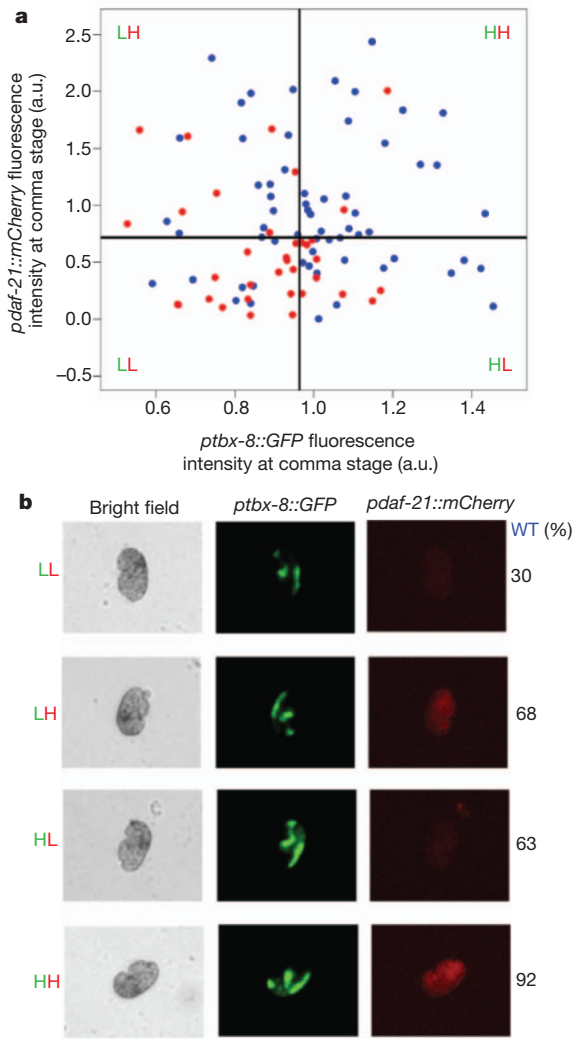


Figure 4 | Simultaneous quantification of inter-individual variation in two buffering systems accurately predicts the outcome of a mutation. **a**, Expression levels of *tbx-8* and *daf-21* reporters in individual *tbx-9(ok2473)* mutant embryos at the comma stage of development are not correlated ($r = 0.061$, $P = 0.56$). Each point represents one individual embryo. Embryos that hatched into phenotypically WT larvae are blue; embryos that hatched with a morphological defect are red. The median values for each reporter are indicated (black lines); embryos with expression above or below the median are defined as high (H) and low (L), respectively. **b**, Percentage of embryos from each expression class hatching with normal morphology. Representative bright-field, green and red fluorescence images are shown. Prediction scores for different time points and measures are given in Supplementary Table 3.

©2011 Macmillan Publishers Limited. All rights reserved

early on in development (Fig. 3d, e, $P = 1.9 \times 10^{-3}$, see Supplementary Table 1). The *daf-21* reporter also predicted *tbx-8*(*ok656*) mutation outcome (Supplementary Figs 15 and 16). Higher chaperone expression during early embryonic development therefore predicts a reduced effect of the inherited mutation.

We next constructed a strain allowing us to quantify simultaneously inter-individual variation in the expression of both buffering systems. Expression from the *tbx-8* and *daf-21* reporters does not correlate across individuals (Fig. 4a, $r = 0.061$, $P = 0.56$), showing that the two buffering systems vary independently. We retrospectively divided *tbx-9*(*ok2473*) mutation embryos into four approximately equally populated groups, depending upon whether their expression was above or below the median for each of the reporters, and examined the proportion of individuals hatching as WT larvae in each group (Fig. 4b). Whereas 30% of embryos with below-median expression of both reporters hatched without an abnormal phenotype ('LL' embryos, Fig. 4a, b, $P = 1.7 \times 10^{-5}$), 68% and 63% of embryos with above-median expression of a single reporter (*daf-21* or *tbx-8*, respectively) showed no morphological defect (Fig. 4b). Strikingly, 92% of embryos with above median expression of both reporters hatched as phenotypically WT larvae ('HH' embryos, Fig. 4a, b). That is, non-genetic variation in two buffering systems can be almost completely epistatic[22] to the *tbx-9* mutation. Receiver operating characteristic curve analysis confirmed the independent contribution of both reporters to phenotypic predictions (Supplementary Fig. 17).

In summary, we have shown here that incomplete penetrance is not just a direct consequence of failure caused by mutations[13,14], but that organisms have plastic compensatory responses[19] in both specific and more promiscuous genetic interaction partners that vary among individuals. It is the combination of this variation that determines the outcome of each mutation in each individual.

Inherited mutations also frequently have variable consequences in other species, including in human disease[2]. Based on our findings in *C. elegans*, we propose that incomplete genetic compensation may play a role in human disease, influencing the outcome of inherited polymorphisms in each individual.

METHODS SUMMARY

The expression from individual developing embryos was quantified using customized imaging and analysis protocols. Embryos were released from adult worms by dissection and sorted according to their developmental stage. Phenotypes were scored at hatching. Reporter constructs were generated by microinjection or bombardment.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 February; accepted 21 October 2011.

1. Badano, J. L. & Katsanis, N. Beyond Mendel: an evolving view of human genetic disease transmission. *Nature Rev. Genet.* **3,** 779–789 (2002).
2. Baranzini, S. E. *et al.* Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature* **464,** 1351–1356 (2010).
3. Horvitz, H. R. & Sulston, J. E. Isolation and genetic characterization of cell-lineage mutants of the nematode *Caenorhabditis elegans*. *Genetics* **96,** 435–454 (1980).
4. Gartner, K. A third component causing random variability beside environment and genotype. A reason for the limited success of a 30 year long effort to standardize laboratory animals? *Lab. Anim.* **24,** 71–77 (1990).
5. Lehner, B., Crombie, C., Tischler, J., Fortunato, A. & Fraser, A. G. Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nature Genet.* **38,** 896–903 (2006).
6. Costanzo, M. *et al.* The genetic landscape of a cell. *Science* **327,** 425–431 (2010).
7. Nowak, M. A., Boerlijst, M. C., Cooke, J. & Smith, J. M. Evolution of genetic redundancy. *Nature* **388,** 167–171 (1997).
8. Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396,** 336–342 (1998).
9. Queitsch, C., Sangster, T. A. & Lindquist, S. Hsp90 as a capacitor of phenotypic variation. *Nature* **417,** 618–624 (2002).
10. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297,** 1183–1186 (2002).
11. Wernet, M. F. *et al.* Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature* **440,** 174–180 (2006).
12. Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E. & Huang, S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453,** 544–547 (2008).
13. Raj, A., Rifkin, S. A., Andersen, E. & van Oudenaarden, A. Variability in gene expression underlies incomplete penetrance. *Nature* **463,** 913–918 (2010).
14. Eldar, A. *et al.* Partial penetrance facilitates developmental evolution in bacteria. *Nature* **460,** 510–514 (2009).
15. Lehner, B. Genes confer similar robustness to environmental, stochastic, and genetic perturbations in yeast. *PLoS ONE* **5,** e9035 (2010).
16. Andachi, Y. *Caenorhabditis elegans* T-box genes *tbx-9* and *tbx-8* are required for formation of hypodermis and body-wall muscle in embryogenesis. *Genes Cells* **9,** 331–344 (2004).
17. Pocock, R., Ahringer, J., Mitsch, M., Maxwell, S. & Woollard, A. A regulatory network of T-box genes and the even-skipped homologue vab-7 controls patterning and morphogenesis in *C. elegans*. *Development* **131,** 2373–2385 (2004).
18. Baugh, L. R. *et al.* Synthetic lethal analysis of *Caenorhabditis elegans* posterior embryonic patterning genes identifies conserved genetic interactions. *Genome Biol.* **6,** R45 (2005).
19. Kafri, R., Bar-Even, A. & Pilpel, Y. Transcription control reprogramming in genetic backup circuits. *Nature Genet.* **37,** 295–299 (2005).
20. DeLuna, A., Springer, M., Kirschner, M. W. & Kishony, R. Need-based up-regulation of protein levels in response to deletion of their duplicate genes. *PLoS Biol.* **8,** e1000347 (2010).
21. Koh, K. & Rothman, J. H. ELT-5 and ELT-6 are required continuously to regulate epidermal seam cell differentiation and cell fusion in *C. elegans*. *Development* **128,** 2867–2880 (2001).
22. Bateson, W. *Mendel's Principles of Heredity* (Cambridge Univ. Press, 1909).
23. Ow, M. C. *et al.* The FLYWCH transcription factors FLH-1, FLH-2, and FLH-3 repress embryonic expression of microRNA genes in *C. elegans*. *Genes Dev.* **22,** 2520–2534 (2008).
24. Vavouri, T., Semple, J. I. & Lehner, B. Widespread conservation of genetic redundancy during a billion years of eukaryotic evolution. *Trends Genet.* **24,** 485–488 (2008).
25. Tischler, J., Lehner, B., Chen, N. & Fraser, A. G. Combinatorial RNA interference in *Caenorhabditis elegans* reveals that redundancy between gene duplicates can be maintained for more than 80 million years of evolution. *Genome Biol.* **7,** R69 (2006).
26. Lehner, B. Conflict between noise and plasticity in yeast. *PLoS Genet.* **6,** e1001185 (2010).
27. Waddington, C. H. Canalization of development and the inheritance of acquired characters. *Nature* **150,** 563–565 (1942).
28. Bobula, J. *et al.* Why molecular chaperones buffer mutational damage: a case study with a yeast Hsp40/70 system. *Genetics* **174,** 937–944 (2006).
29. Van Dyk, T. K. G. A. LaRossa RA. Demonstration by genetic suppression of interaction of GroE products with many proteins. *Nature* **342,** 451–453 (1989).
30. Tokuriki, N. & Tawfik, D. S. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* **459,** 668–673 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was funded by grants from the European Research Council, Institució Catalana de Recerca i Estudis Avançats, Ministerio de Ciencia e Innovación Plan Nacional BFU2008-00365, Agència de Gestió d'juts Universitaris i de Recerca, ERASysBio+, the European Molecular Biology Organization Young Investigator Programme, the EMBL-CRG Systems Biology Program, by a Formación de Personal Investigador–Ministerio de Ciencia e Innovación fellowship to A.B. and by a Beatriu de Pinós Fellowship to M.O.C. We thank I. Hope, V. Ambros and S. Kim for providing strains. Additional strains were obtained from the Caenorhabditis Genetics Center, which is funded by the National Institutes of Health National Center for Research Resources. We thank T. Zimmermann and R. García from the CRG Advanced Light Microscopy Unit for advice and assistance, J. Miwa and Y. Yamaguchi for 608F antibody, J. Semple for providing complementary DNA clones, A. Marchetti and R. García-Verdugo for technical assistance, J. Tischler and C. Kiel for advice on single-molecule fluorescence *in situ* hybridization and western blotting, respectively, and L. Serrano, M. Isalan and J. Semple for comments on the manuscript.

Author Contributions A.B. performed all experiments, developed the method and analysed the data; M.O.C. demonstrated that increased chaperone activity can suppress mutation outcome in *C. elegans*; A.B. and B.L. designed experiments, conceived the model and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.L. (ben.lehner@crg.eu).

©2011 Macmillan Publishers Limited. All rights reserved

METHODS

Expression profiling: image acquisition. Synchronous populations of *tbx-9*(*ok2473*) animals were obtained by treating worms with sodium hypochlorite and then allowing embryos to develop at 25 °C. For a typical time-lapse, around 30–50 gravid worms were picked and then washed three times with M9 buffer in a watch glass. Worms were then transferred to a 50 mm glass-bottomed culture dish (MatTek Corporation). Adults were dissected and released embryos were manually sorted. In a period of time not exceeding 15 min, 4- to 12-cell-stage embryos were collected and carefully placed with an eyelash-pick into the centre of the dish; the rest of embryos and debris were removed using a pipette. A cover slip was placed on top of the worms and sealed with paraffin. Fluorescence and bright-field images in multiple positions were acquired with a ×10 Plan-NEOFLUAR 0.3 NA objective on a Zeiss Cell Observer HS system consisting of an inverted microscope (AxioObserver.Z1) equipped with an automated stage, a C9100-13 IMAG-EM Dual Mode EM-CCD camera (Hamamatsu) and a Sutter DG-4 fast switching xenon light source (Sutter Instrument Company). Temperature control was achieved using a custom microscope incubator box and temperature control device developed by the European Molecular Biology Laboratory (http://www.embl-em.de) coupled to a Unichiller cooling unit (Huber Kältemaschinenbau) for temperature control below room temperature.

At the start of imaging, approximately 25–30 min had elapsed since the dissection of worms. During the first 20 min of the time-lapse no fluorescence illumination was applied, to avoid phototoxicity as early embryos are very sensitive to illumination. Subsequently bright-field and fluorescence images were obtained every 10 min for 8 h, followed by just bright-field images collected every 2 min for 4 more hours or until most of the worms had hatched. The increase in bright-field frequency facilitated the phenotypic score of hatching worms. Exposure times and frequency were adjusted to maximize the signal-to-noise ratio while avoiding phototoxicity (evaluated as approximately 100% hatching of WT (N2) embryos).

The imaging conditions for the *daf-21*, *let-858*, *elt-5* and *flh-2* reporters in a *tbx-9* mutant background were exactly the same as for the *tbx-8* reporter. In the other genetic backgrounds exposure times and acquisition characteristics were adjusted according to the characteristics of the reporter construct. The *tbx-9* reporter was expressed earlier than the *tbx-8* reporter and at lower levels, and the exposure time was increased 2.5× to compensate for this, while reducing the frequency of acquisition to once every 25 min. *The flh-2* reporter was also expressed at lower levels, but the induction was later in development. For this reporter the exposure time was increased (2.5× compared with that for the *tbx-8* reporter), fluorescence acquisition was initiated 1 h later than bright field, but images were still obtained once every 10 min.

When comparing the expression of a particular reporter in two different genetic backgrounds, embryo isolation was performed in 10 min, one strain immediately after the other (total time of 20 min), and image acquisition was performed simultaneously using the conditions previously established for each strain.

Expression profiling: quantification of gene expression. Total fluorescence from each embryo was quantified using ImageJ[31] (http://rsb.info.nih.gov/ij/). First, we performed a global background correction of all images by subtracting an image taken in a region of no fluorescence for each time point. For the *pelt-5::mCherry* reporter data set, an additional flat field correction step was necessary. Background-subtracted images were divided by the homogeneous fluorescence reference to correct for non-homogeneous illumination, and the resulting images used for quantification[32]. The homogeneous fluorescent reference was created by imaging a drop of 0.05 mg ml^{-1} Nile Red solution in M9 buffer. All embryos were visually identified and tagged so that no embryo was counted twice. A region of interest (ROI) was manually drawn around each embryo and an equally sized ROI drawn in the background vicinity; finally, the total fluorescence from each embryo was calculated as the integrated density difference between the embryo containing ROI and the local background ROI. For the particular case of the *flh-2* reporter, at the 100-cell stage most of the reporter signal came from a very restricted area of the embryo. Because embryo background fluorescence was masking this specific source of expression, we quantified it by first drawing an equally sized circle-shaped ROI in all embryos around the signal source and measuring the integrated density. This quantified expression was used for *flh-1* phenotypic prediction.

All embryos were staged using three developmental landmarks: (1) number of cells at beginning of time-lapse, (2) elapsed time until comma stage and (3) elapsed time until twofold-stage. Only those embryos starting the time-lapse at the approximately 6- to 15-cell stage and reaching the next development landmarks with a delay of less than or equal to 30 min were considered for further analysis. In a typical time-lapse, around 10–15 embryos fulfilled those requirements. Data for each reporter–mutant pair prediction included embryos from at least four independent experiments. To correct for day-to-day technical variation (Supplementary Fig. 18), the expression of each embryo was normalized using the mean expression of the population (or populations when comparing the expression of different genotypes) for each time-lapse. For visualization, five time-point smoothing was used for 10 min acquisition frequency recordings, and three time-point smoothing used when the frequency was 25 min.

Data analysis, statistics and figures were generated using R (http://www.r-project.org). Receiver operating characteristic curves for dual reporter predictions were constructed using the ROCR package[33]. Expression levels of *tbx-8* and *daf-21* at comma stage were normalized by subtracting the mean and dividing by the standard deviation of each population. The sum of the normalized values for each embryo was used to calculate the receiver operating characteristic curve of the joint contribution.

Phenotypes. The phenotype of each embryo was scored as WT or abnormal by visual inspection of the hatching morphology. Any abnormal morphology was defined as non-WT. To study the prediction of mutation expressivity, *tbx-9*(*ok2473*) *animals* that hatched with an abnormal phenotype were further classified into two classes. 'Mild' abnormality indicated individuals with a characteristic posterior depression but that conserve the general body structure. 'Severe' embryos were those with severe malformations along most of the posterior axis or the entire body (Supplementary Fig. 19).

Nematode strains and growth conditions. Worms were grown at 20 °C on NGM plates using *Escherichia coli* OP50 as a food source unless otherwise noted. The strains used in and constructed for this study are listed in Supplementary Table 2. Wild type was Bristol N2 (ref. 34). Reporter strains *ptbx-8::GFP* (UL2520) and *pelt-5::GFP* (SD1434) were provided by I. Hope and S. Kim, respectively. The strain VT1343 (*flh-1*(*bc374*)) was a gift from V. Ambros. Some deletion mutations used in this work were provided by the *C. elegans* Gene Knockout Consortium.

We optimized worm growth and embryo imaging conditions for each mutant, to obtain a penetrance with a reasonable representation of embryos hatching with both WT and morphologically abnormal phenotypes. The *tbx-9*(*ok2473*) mutants were maintained at 25 °C and embryos were allowed to develop at 27.5 °C to give a penetrance of approximately 50%; *tbx-8*(*ok656*) were grown at 16 °C and embryos allowed to develop at 20 °C; *flh-1*(*bc374*) mutants were grown at 20 °C and embryos allowed to develop at 26 °C. For *flh-2*(*bc375*) mutants, a growth condition was not found where the penetrance was high enough to be usefully studied.

Reporter gene constructs. We generated transcriptional reporters for the genes *tbx-9*, *flh-2* and *daf-21*. Promoter regions of these genes were amplified from N2 genomic DNA by PCR using the following primer pairs: FW-ptbx-9, 5′-TTGGGTTCAGATAACAATTTGG-3′; RV-ptbx-9, 5′-ATTTTTTGTCTGAAAACGTTAAAATA-3′; FW-pflh-2, 5′-GCGCTTCTCGTGGGCTCT-3′; RV-pflh-2, 5′-ATACAGGCGGTCTGAAAA-3′; FW-pdaf-21, 5′-CGAAACGGTCGAATTTCATAA-3′; RV-pdaf-21, 5′-ATGGTTCTGGAAAAATATCAATTA-3′. The lengths of the amplified regions were 1.5 kb, 1.5 kb and 1.9 kb, respectively. The promoters were cloned into the MultiSite Gateway Entry vector pENTR 5′-TOPO following the manufacturer's instructions (Invitrogen) and then transferred by three-fragment MultiSite Gateway Pro LR reaction (Invitrogen) into the pCFJ150-pDESTttTi5605[R4-R3] destination vector. Middle Entry vectors used were pENTRwGFP or pENTRmCherry; 3′ Entry vector was pCM5.37 containing the *unc-54* 3′ untranslated region (UTR). Final constructs were confirmed by PCR amplification and sequencing.

Strain construction. Strains carrying an extrachromosomal array expressing *ptbx-8::TBX-8::unc-54* 3′ UTR or *ptbx-9::TBX-9::unc-54* 3′ UTR were generated by microinjection as previously described[35] using *pmyo-2::mCherry* as a co-injection marker. The transcriptional reporter *ptbx-9::GFP* was generated by microinjection using *pmyo-2::mCherry* as a co-injection marker, followed by integration using the ultraviolet irradiation method[36]. The transcriptional reporters *pdaf-21::mCherry*, *pdaf-21::GFP*, *phsp-4::mCherry*, *ptbx-8::mCherry* and *pflh-2::GFP* were generated by bombardment in an *unc-119*(*ed3*) background[37,38]. All transgenic strains were backcrossed at least four times. No transgenic strains displayed abnormal phenotypes.

Quantitative PCR. Worms carrying the *pdaf-21*::mCherry transgene were synchronized by hypochlorite treatment, plated as L1 larvae and allowed to develop for 24 h at 25 °C. Larvae were then sorted by hand using a MVX10 Macro Zoom fluorescence microscope (Olympus) into 'high' and 'low' groups based on their mCherry fluorescence levels. Each group had approximately 20 larvae and sorting was performed in triplicate. Worms were washed in M9 and transferred to 1.5 ml Eppendorf tubes containing 200 µl of lysis Buffer (50 mM KCl, 10 mM Tris-HCl (pH 8), 2.5 nM $MgCl_2$, 0.45% Nonidet P-40, 0.45% Tween20) supplemented with 20 µg proteinase K and 20 µg RNase A. Tubes were frozen in liquid nitrogen, followed by incubation at 65 °C for 1 h and 95 °C for 30 min. Genomic DNA was extracted by adding 500 µl Phenol:Chloroform:Isoamyl Alcohol 25:24:1 (Sigma) and precipitated using 600 µl of ice cold ethanol and 50 µl of 3 M NaAc. Quantitative PCR was performed on a LightCycler 480 machine with SYBR green detection (Roche). The transgene was amplified with primers targeting the mCherry coding region: FW-mCherry, 5′-CTACGACGCTGAGGTCAAGA-3′; RV-mCherrry, 5′-CGATGGTGTAGTCCTCGTTG-3′. Transgene levels were normalized by the mean level of two other loci in the genome using

©2011 Macmillan Publishers Limited. All rights reserved

the following primers: FW-daf-21ORF, 5′-TCCAATGACTGGGAAGATCA-3′; RV-daf-21ORF, 5′-CGAACGTAGAGCTTGATGGA-3′; FW-csq-1, 5′-AACTG AGGTTCTGACCGAGAAG-3′; RW-csq-1, 5′-TACTGGTCAAGCTCTGAGT CGTC-3′.

Western blotting. To test whether expression of the *pdaf-21::mCherry* predicted differences in the levels of DAF-21 protein expressed from the endogenous gene, *pdaf-21::mCherry; glp-1*(*e2141*) animals were synchronized as L1 larvae by hypochlorite treatment and starvation at 20 °C overnight. Worms were transferred to plates and allowed to develop for 24 h at 25 °C (*glp-1*(*e2141*) mutation restrictive temperature; this mutation prevents the development of the germ line; *the pdaf-21::mCherry* reporter does not express in germline cells). Larvae were then sorted using an MVX10 Macro Zoom fluorescence microscope (Olympus) into 'high' and 'low' groups based on their mCherry fluorescence. Each group had approximately 30 larvae and sorting was performed in triplicate. Worms were washed in M9 buffer, frozen in liquid nitrogen, dissolved in 30 µl of SDS loading buffer, sonicated for 5 min at 4 °C and then boiled at 95 °C for 5 min. Samples were loaded into a 12.5% SDS polyacrylamide gel and transferred to a nitrocellulose membrane. Primary antibodies were incubated overnight at 4 °C. The following antibodies and dilutions were used: rabbit polyclonal anti-β-actin (ab8227) (Abcam) 1:1,000; rabbit polyclonal anti-DsRed (ClonTech) 1:1,000; mouse monoclonal anti-DAF-21 608F hybridoma supernatant[39] (a gift from J. Miwa and Y. Yamaguchi) 1:40; anti-rabbit IgG peroxidase conjugate (Jackson ImmunoResearch) 1:40,000 and anti-mouse IgG peroxidase conjugate (Sigma-Aldrich) 1:8,000. Quantification of DAF-21 levels in N2 WT and *daf-21*(+/−) was performed in a similar fashion; 30 L4 larvae of each genotype were picked in triplicate. We detected peroxidase activity with SuperSignal West Femto Substrate (Thermo Scientific) and blot imaging was performed using a Fujifilm LAS-3000 luminescent image analyser with exposures times in the lineal range of detection. Finally, band intensities were quantified using ImageJ.

Single molecule fluorescence *in situ* hybridization. We followed the protocol of Raj *et al.*[40] with the following modifications. Probes against *tbx-8* mRNA were generated with the online tool available at www.singlemoleculefish.com and synthesized by Biosearch Technologies. Probes were coupled to CAL Fluor Red 590 fluorophore. The formamide concentration in hybridization and wash buffer was 20% and the total concentration of pooled probes was 25 nM. Hybridization was performed overnight at 30 °C. We imaged embryos using an oil immersion ×100 objective on a Leica DMI4000 inverted microscope equipped with an Evolve 512 EMCCD camera (Photometrics) and a Lumen 200 metal arc lamp (Prior Scientific). We collected stacks of typically 25 images spaced by 0.3 µm for each embryo. Stacks were processed applying a Laplacian of Gaussian filter with the LoG3D plugin[41] for ImageJ. Then dimensionality was reduced by applying a Z-maximum intensity projection (the final image contains the maximum value over all images in the stack at the particular pixel location), and a single threshold was applied to each embryo by visual inspection as recommended. The total number of particles was counted with the 'Analyze particles' ImageJ tool. Embryos were staged by counting the total number of nuclei stained with 4′,6-diamidino-2-phenylindole (DAPI).

Rescue experiments. Worms mutant for *tbx-9* were crossed with a line carrying an extrachromosomal array overexpressing TBX-8 protein and a *pmyo-2::mCherry* reporter as a marker. Adult gravid worms growing at 25 °C were manually sorted under an MVX10 Macro Zoom fluorescence microscope (Olympus) into two groups, one carrying the fluorescent marker, the other lacking it. This latter group served as an internal negative control. Extrachromosomal array transmission was around approximately 50%, meaning that not all embryos in the overexpression group actually carried the transgene, so rescue scores are probably an underestimation. Four- to 12-cell stage embryos were collected in 50 mm glass-bottomed dishes following the same protocol as for time-lapses, and incubated at 27.5 °C. Phenotypes of worms were scored 15 h later. In the case of *tbx-8* mutation rescue, worms were grown at 16 °C and incubated at 20 °C. Control experiments were performed crossing the *tbx-9*(*ok2473*) and *tbx-8*(*ok656*) mutations to a strain carrying the extrachromosomal array *wwEx26*[*pmir-42-44::GFP+unc-119*(+)]. This transcriptional reporter is expressed at high levels during early and late development, and its transmission is approximately 95%. Rescue experiments were performed at least in triplicate for each pair. Finally, repeats were pooled and significance was tested using Fisher's exact test.

31. Abramoff, M. D., Magelhaes, P. J. & Ram, S. J. Image processing with ImageJ. *Biophoton. Int.* **11,** 36–42 (2004).
32. Wolf, D. E., Samarasekera, C. & Swedlow, J. R. Quantitative analysis of digital microscope images. *Methods Cell Biol.* **81,** 365–396 (2007).
33. Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. ROCR: visualizing classifier performance in R. *Bioinformatics* **21,** 3940–3941 (2005).
34. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77,** 71–94 (1974).
35. Kelly, W. G., Xu, S., Montgomery, M. K. & Fire, A. Distinct requirements for somatic and germline expression of a generally expressed *Caernorhabditis elegans* gene. *Genetics* **146,** 227–238 (1997).
36. Mitani, S. Genetic regulation of mec-3 gene expression implicated in the specification of the mechanosensory neuron cell types in *Caenorhabditis elegans*. *Dev. Growth Differ.* **37,** 551–557 (2003).
37. Praitis, V., Casey, E., Collar, D. & Austin, J. Creation of low-copy integrated transgenic lines in *Caenorhabditis elegans*. *Genetics* **157,** 1217–1226 (2001).
38. Wilm, T., Demel, P., Koop, H. U., Schnabel, H. & Schnabel, R. Ballistic transformation of *Caenorhabditis elegans*. *Gene* **229,** 31–35 (1999).
39. Yamaguchi, Y., Murakami, K., Furusawa, M. & Miwa, J. Germline-specific antigens identified by monoclonal antibodies in the nematode *Caenorhabditis elegans*. *Dev. Growth Differ.* **25,** 121–131 (1983).
40. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nature Methods* **5,** 877–879 (2008).
41. Sage, D., Neumann, F. R., Hediger, F., Gasser, S. M. & Unser, M. Automatic tracking of individual fluorescence particles: application to the study of chromosome dynamics. *IEEE Trans. Image Process.* **14,** 1372–1383 (2005).

©2011 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature10575

Mapping intact protein isoforms in discovery mode using top-down proteomics

John C. Tran[1,2], Leonid Zamdborg[1], Dorothy R. Ahlf[1,2], Ji Eun Lee[1,3], Adam D. Catherman[1,2], Kenneth R. Durbin[1,2], Jeremiah D. Tipton[2], Adaikkalam Vellaichamy[1]†, John F. Kellie[1,2], Mingxi Li[1,2], Cong Wu[1], Steve M. M. Sweet[1,2], Bryan P. Early[1,2], Nertila Siuti[1]†, Richard D. LeDuc[4], Philip D. Compton[2], Paul M. Thomas[1,2] & Neil L. Kelleher[1,2]

A full description of the human proteome relies on the challenging task of detecting mature and changing forms of protein molecules in the body. Large-scale proteome analysis[1] has routinely involved digesting intact proteins followed by inferred protein identification using mass spectrometry[2]. This 'bottom-up' process affords a high number of identifications (not always unique to a single gene). However, complications arise from incomplete or ambiguous[2] characterization of alternative splice forms, diverse modifications (for example, acetylation and methylation) and endogenous protein cleavages, especially when combinations of these create complex patterns of intact protein isoforms and species[3]. 'Top-down' interrogation of whole proteins can overcome these problems for individual proteins[4,5], but has not been achieved on a proteome scale owing to the lack of intact protein fractionation methods that are well integrated with tandem mass spectrometry. Here we show, using a new four-dimensional separation system, identification of 1,043 gene products from human cells that are dispersed into more than 3,000 protein species created by post-translational modification (PTM), RNA splicing and proteolysis. The overall system produced greater than 20-fold increases in both separation power and proteome coverage, enabling the identification of proteins up to 105 kDa and those with up to 11 transmembrane helices. Many previously undetected isoforms of endogenous human proteins were mapped, including changes in multiply modified species in response to accelerated cellular ageing (senescence) induced by DNA damage. Integrated with the latest version of the Swiss-Prot database[6], the data provide precise correlations to individual genes and proof-of-concept for large-scale interrogation of whole protein molecules. The technology promises to improve the link between proteomics data and complex phenotypes in basic biology and disease research[7].

Effective fractionation[8–10] is critical for sample handling before mass-spectrometry-based proteomics. So far, no fractionation procedure for intact proteins can match the resolution of two-dimensional gel electrophoresis (two-dimensional gels). Here we use a liquid phase alternative to two-dimensional gels that bypasses both their low recovery and extensive workup steps before mass spectrometry[11]. This procedure for two-dimensional liquid electrophoresis[12] comprises solution isoelectric focusing (sIEF) followed by gel-eluted liquid fraction entrapment electrophoresis (GELFrEE)[13] for fractionation by protein isoelectric point and size, respectively (Fig. 1a, b). Combining these with nanocapillary liquid chromatography and mass spectrometry (LC–MS) (Fig. 1c) for both low[14] and high molecular mass proteins[15] results in an overall four-dimensional separation of whole protein molecules before ion fragmentation by tandem mass spectrometry and protein identification.

Using the four-dimensional platform described above, we generated a quasi-two-dimensional gel perspective of the human proteome with extremely high molecular detail (Fig. 2a) from individual replicate

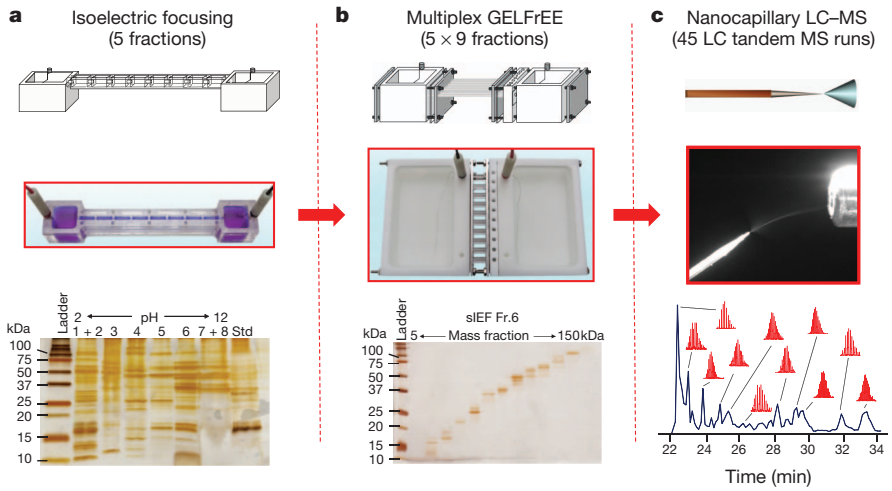


Figure 1 | The four-dimensional platform for high-resolution fractionation of protein molecules. Schematics (top) and photographs (middle) are shown for (**a**) a custom device for sIEF, (**b**) a custom device for multiplexed GELFrEE and (**c**) RPLC coupled to mass spectrometry. Representative one-dimensional gels of fractions collected from the two electrophoretic devices are shown below their pictures; note the resolution attainable at the level of intact proteins. The combined resolution of RPLC with Fourier-transform mass spectrometry is depicted by the chromatogram along with selected isotopic distributions for protein ions measured during the run.

[1]Departments of Chemistry and Biochemistry, and the Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. [2]Departments of Chemistry and Molecular Biosciences and the Feinberg School of Medicine, Northwestern University, Evanston, Illinois 60208, USA. [3]Doping Control Center and Center for Theragnosis, Korea Institute of Science and Technology, Seoul, South Korea. [4]Department of Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA. †Present addresses: Department of Nanomedicine, The Methodist Hospital Research Institute, Houston, Texas 77030, USA (A.V.); Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA (N.S.).

©2011 Macmillan Publishers Limited. All rights reserved

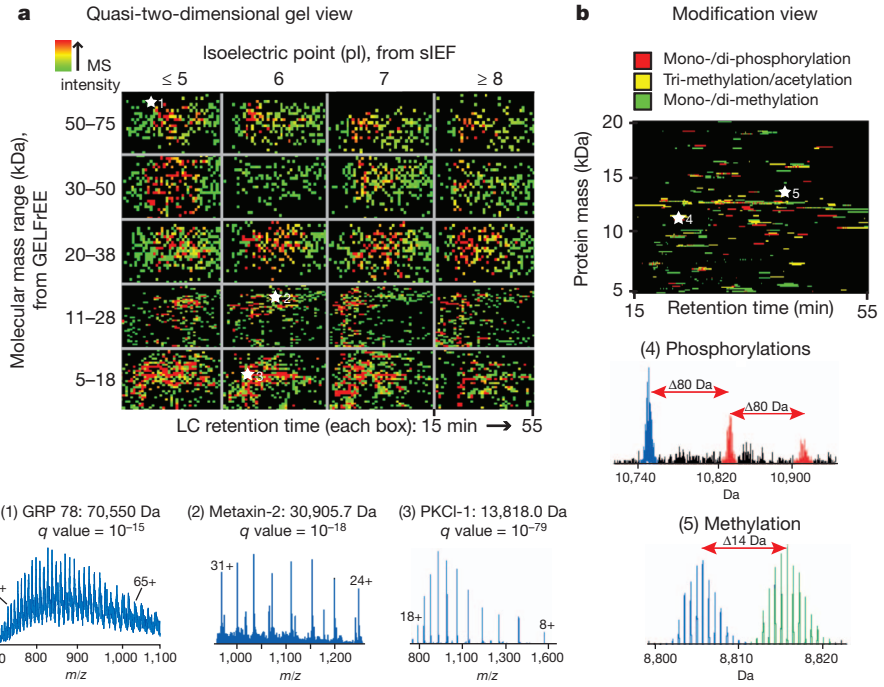


Figure 2 | Two visual representations of proteome-scale runs. **a**, The heat map is generated from combined four-dimensional runs of nuclear and cytosolic extracts. Intact mass and isoelectric point values are indicated on the *y*- and *x*-axes, respectively. Each box in the grid displays total ion chromatograms from LC runs of two-dimensional liquid electrophoresis fractions plotted as time versus neutral intact mass. The mass spectrometry intensity is indicated by colour (legend on top left). Representative precursor scans were extracted from the heat map for electrospray ionization–mass spectrometry spectra of high (1), medium (2) and low (3) mass proteins, along with their identifications from online fragmentation (insets at bottom). **b**, Plot created from selective display of protein pairs with mass differences consistent with acetylation (yellow), phosphorylation (red) and methylation (green), with three and two protein species shown as examples in insets (4) and (5), respectively.

analyses of nuclear and cytosolic extracts of HeLa S3 cells (Supplementary Fig. 1). In discovery mode, the IEF–GELFrEE-nanocapillary liquid chromatography platform used 0.5–1 mg of input protein and provided a peak capacity of well over 2,000 for separation of protein molecules in solution. Considering the separation power of the mass spectrometer, the peak capacity of the four-dimensional system is greater than 100,000 for proteins below approximately 25 kDa (Supplementary Information). This is 20-fold higher than the peak capacity for high-resolution two-dimensional gels (less than 5,000). Identification and characterization of isoforms were achieved using fragmentation data acquired with less than 10 part-per-million mass accuracy for searching databases with highly annotated primary sequences[16]. Using tailored software[17], we overcame the 'protein inference problem' where identification ambiguity results when isoforms (for example, from members of a gene family or alternative splicing) produce many identical tryptic peptides[2,18]. The databases and search engine used here are fully compatible with the UniProt flat file format and enable a deep consideration of known PTMs, alternative splice variants, polymorphisms, endogenous proteolysis and diverse combinations of all these sources of molecular variation at the protein level[16]. Together with the careful curation of the Swiss-Prot database[6], the result is an informatics framework that maps each given protein identification to a single gene (except in rare cases like ubiquitin, where multiple genes can produce the identical sequence). Extended details on statistical analysis are provided in the Methods section.

A total of 1,043 proteins were identified with unique Swiss-Prot accession numbers in this study (Supplementary Table 1). These identifications originate from 1,045 human genes, 77% of whose protein products displayed amino (N)-terminal acetylation. The distribution of *q* values, which indicates the confidence of protein identifications (see Methods), is shown in Fig. 3c. This level of proteome coverage represents the most comprehensive implementation of top-down mass spectrometry so far, with an approximate tenfold increase in identifications of intact proteins for any microbial system[19–21] and a greater than 20-fold increase over any previous work in mammalian cells[14,22] (Fig. 3a). In addition, fragmentation evidence for 3,093 protein isoforms/species was captured in this initial report (Supplementary Table 1), with PTMs detected as follows: 645 phosphorylations, 538 lysine acetylations, 158 methylations, 19 lipid/terpenes and 5 hypusines. Over 400 species were attributed to core histones alone. Comparisons of predicted protein hydrophobicity and isoelectric point showed minimal bias compared with that expected for the human proteome (Supplementary Fig. 2).

Using an orthogonal method to detect PTMs based on intact mass values[17], we detected pairs of protein species showing characteristic mass differences (Fig. 2b). For proteins less than 20 kDa, 225 pairs showed mass differences consistent within 0.05 Da with mono-methylation, 185 with di-methylation and 122 with tri-methylation/acetylation. Other mass differences revealed 87 cases consistent with double acetylation, 140 with mono-phosphorylation and 100 with di-phosphorylation events (Fig. 2b). Using this set of mass differences on the entire HeLa data set for all isotopically resolved proteins, a total of 2,130 such mass shifts were found.

Complete characterization of a protein requires the theoretical and experimental mass values to match within error. For the 1,043 proteins identified, 431 and 331 were identified with intact mass information from either isotope spacings or deconvolution of charge states, respectively. Of these data, 54% of the isotopically resolved proteins matched the species identified from the database within 2 Da (Supplementary Fig. 3a). Likewise, 130 of 331 of the masses determined by deconvolution were manually determined to be of high quality and 51% of these matched within 200 Da (Supplementary Fig. 3b). The protein species outside these windows are clearly identified by fragmentation, but harbour unexplained mass discrepancies (Δm) at this time. The complete explanation of Δm in the human proteome motivates future refinements in data acquisition to obtain enough tandem mass spectrometry information on all the protein isoforms/species.

Major functional differences can exist among protein isoforms in a family, making their precise identification a major boost in the

©2011 Macmillan Publishers Limited. All rights reserved

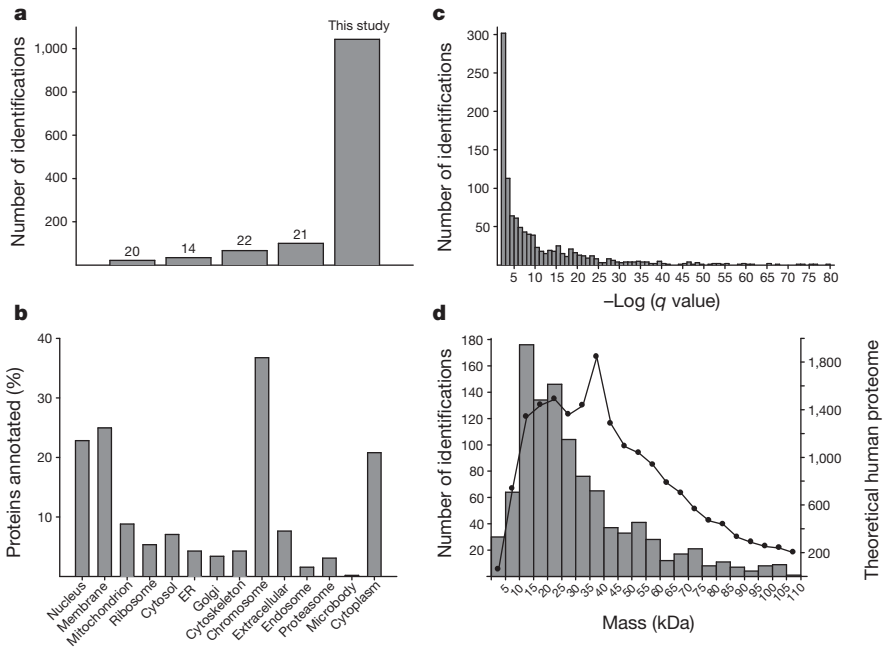


Figure 3 | Proteome analysis metrics associated with this study. **a**, Graph showing the striking increase in identifications from previous studies achieved in archaeal, bacterial, yeast or human systems. References cited are indicated above each bar. **b**, A gene ontology analysis for the identifications in this study. **c**, Histogram showing the distribution of *q* values for the identified proteins. **d**, Plot showing the molecular mass distribution for the unique identifications obtained. The line graph depicts the theoretical molecular mass distribution for the human proteome (Swiss-Prot, *Homo sapiens*, 20,223 entries).

information content of proteomic analyses in higher eukaryotes. An intact protein mass and matching fragment ions from both termini are usually sufficient to accomplish a gene-specific identification[4,17]. Here, 9 of the approximately 15 isoforms of histone H2A were fully characterized in an automated fashion despite their greater than 95% sequence identity (including the H2A.Z and H2A.X variants) with an additional three having Δm greater than 1 Da (H2A type 1-D, 2-C and 2-B). Also identified were nine S100 proteins, several α- and β-tubulins, seven unique isoforms of human keratin (a widely known contaminant in proteomics), MLC20, BTF3 and their related sequences (which are 97% and 81% identical, respectively (Supplementary Figs 4 and 5)), and over 100 isoforms/species from the high mobility group (HMG) family (see, for example, Fig. 4). Significant improvements for top-down proteomics in discovery mode were made for proteins in the 40–110 kDa range (Fig. 3d), including extensive characterization of GRP78, a 70.6-kDa heat-shock protein (more than 12 fragment ions mapping to each terminus, Supplementary Fig. 6), and identification of several proteins greater than 90 kDa, such as P33991 and Q14697 at 97 and 104 kDa, respectively (Supplementary Table 2).

Because the two-dimensional liquid electrophoresis platform makes use of SDS extensively, we anticipated reduced bias against integral membrane proteins. In all, 32% of the 1,043 total identifications from HeLa cells were membrane-associated proteins (GO:0016020), with 62% of these annotated as integral membrane proteins (GO:0016021, Supplementary Table 4). A more focused study of a mitochondrial membrane fraction (see Methods) used chromatographic procedures modified for enhanced separation of membrane proteins. We identified an additional 46 integral membrane proteins (Supplementary Table 3) from a single three-dimensional experiment (no isoelectric focusing). Detailed inspection of the species that eluted from the column during LC–MS revealed proteins with a distribution of 1–11 transmembrane helices (Supplementary Table 3). This shows a broad applicability of this study and will drive further efforts to detect full-length isoforms of membrane proteins[23].

As part of our study of the HeLa proteome, cells were treated with etoposide to elicit the DNA damage response (see Methods), followed by four-dimensional fractionation and top-down tandem mass spectrometry. Using gene ontology (GO) analysis, we annotated all

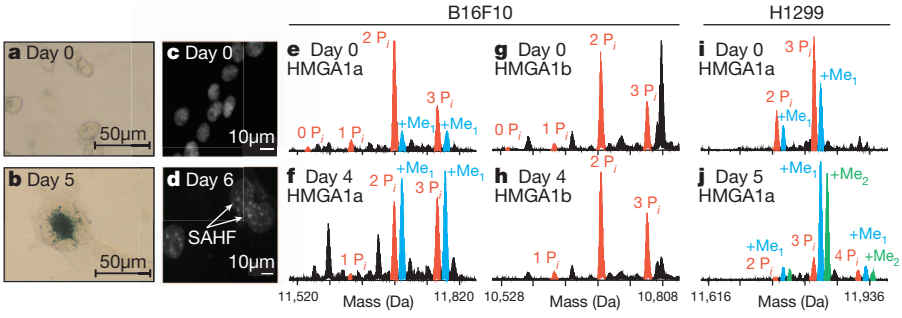


Figure 4 | Monitoring dynamics of HMGA1 isoforms during senescence in B16F10 and H1299 cells. After induction of DNA damage by transient treatment with camptothecin for H1299 cells or etoposide for B16F10, progression of accelerated senescence was monitored by SA-β-Gal (**a**, **b**) or DAPI staining to monitor formation of senescence-associated heterochromatic foci (SAHF) (**c**, **d**) over the specified recovery period. Changes in modification profiles on HMGA1a (**e**, **f**) and HMGA1b (**g**, **h**) from B16F10 showed mild increases in phosphorylation occupancy but a significant increase in methylation levels on multiply phosphorylated species. A more striking increase in both methylation and phosphorylation was observed in senescent H1299 cells (**i**, **j**). No such methylations were observed in the HMGA1b profiles for either cell line.

©2011 Macmillan Publishers Limited. All rights reserved

four-dimensional identifications according to cell compartment (Fig. 3b) or biological process (Supplementary Fig. 7). Many proteins detected were involved in cell cycle regulation and apoptosis, including nine that interact with proliferating cell nuclear antigen during repair of DNA damage (Supplementary Fig. 8). Also, several proteins involved in the Fanconi anaemia pathway were identified including FANCE, RAD51AP1, RAD23B and RPA3, with the last two completely characterized (Supplementary Table 5). Several CDK inhibitors were found, such as $p27^{Kip1}$ (CDKN1B) and $p16^{INK4a}$ (CDKN2A), T53G1 and the protein product from a target gene of p53 (Q9Y2A0, p53-activated protein 1).

Using the three-dimensional fractionation approach (that is, GELFrEE-nanocapillary LC–MS) to readout phosphorylation stoichiometry with high fidelity (Supplementary Information and Supplementary Fig. 9), we monitored 17 phosphoprotein targets across three time points at three different concentrations of etoposide (Supplementary Table 6). We found increases in the occupancy of phosphorylation in H2A.X-pSer139 (γH2A.X) after treatment with 25 or 100 μM etoposide for 1 h (Supplementary Fig. 10). After a 24-h recovery from treatment, a return to basal levels of phosphorylation of γH2A.X was found, consistent with engagement of the DNA-repair machinery[24]. Further, we observed a strong correlation between the phosphorylation stoichiometry of γH2A.X determined by mass spectrometry with the results from immunofluorescence and western blotting run in parallel (Supplementary Fig. 10a–c).

In separate studies we tracked over 2,300 species (from 690 proteins) in H1299 cells (Supplementary Table 7) and 2,300 species (from 708 proteins) in B16F10 melanoma cells (Supplementary Table 8) in the days after a 24-h treatment with camptothecin or 5 h of etoposide, respectively, using only the three-dimensional fractionation approach. After induction of DNA damage, we also monitored the classic hallmarks of stress-induced senescence in H1299 (ref. 25) and B16F10 (ref. 26) over several days (Supplementary Fig. 11a–c), including cell enlargement and formation of senescence-associated heterochromatic foci (Supplementary Fig. 11d–f). Although levels of γH2A.X remained the same as in control cells, a striking upregulation in methylated forms of di- and tri-phosphorylated HMGA1a, but not of its splice variant HMGA1b was observed as both B16F10 and H1299 cells entered stress-induced senescence (Fig. 4 and Supplementary Fig. 11g–l).

Full descriptions of the fragmentation data for two multiply modified species of HMGA1 are presented in Supplementary Fig. 12. In mapping these species, the hierarchy of phosphorylations on HMGA1a was determined for control cells to be Ser 101 and Ser 102 occupied in the $2P_i$ form and evidence for the third site pointing predominantly towards pSer 98. The $3P_i$ and $4P_i$ forms both showed some occupancy for pSer 43 (data not shown), a site only available in the splice region specific to the HMGA1a variant (Supplementary Fig. 12). For day 5 in senescent H1299 cells, the effect on methylation was particularly dramatic, with both the mono- and di-methylated species (also harbouring multiple phosphorylations) reproducibly increased to be greater than 80% of the total signal for species from the *hmga1* gene (Fig. 4 and see Supplementary Fig. 13 for biological replicates). The methylation site was localized precisely to Arg 25 (Supplementary Fig. 12), consistent with previous work on HMGA1 proteins[27]. A similar response for methylated HMGA species has been observed in damaged cancer cells undergoing apoptosis[27,28] but the B16F10 and H1299 cells prepared here were clearly senescent as measured by annexin V staining and fluorescence-activated cell sorting analysis through day 6 (data not shown). As Arg 25 is in the first AT-hook DNA-binding region (residues 21–31), it is possible that the R25me1 and R25me2 marks perturb DNA-kinking and allow HMGA1a to be preferentially incorporated into senescence-associated heterochromatic foci[29] during accelerated cellular senescence. Other changes in bulk chromatin were also notable, such as hypoacetylation on all core histones, increased levels of H3.2K27me2/3, and decreased H3.2K36me3.

The sharp increase in proteome coverage demonstrated here provides a path ahead for interrogating the natural complexity of protein primary structures that exist within human cells and tissues. Because this is the first time top-down proteomics has been achieved at this scale, an early glimpse at the prevalence of uncharacterized mass shifting events has been revealed in the human proteome. With faithful mapping of intact isoforms on a proteomic scale, detecting co-variance in modification patterns will help lay bare the post-translational logic of intracellular signalling. Also, proper speciation of protein molecules offers the promise of increased efficiency for biomarker discovery through stronger correlations between measurements and organismal phenotype (for example, a particular isoform of apolipoprotin C-III and levels of high-density lipoprotein in human blood[7]). Technology for intact protein characterization could also become a central approach to focus an analogous effort to the human genome project—to provide a definitive description of protein molecules present in the human body[30].

METHODS SUMMARY

For large-scale global analysis, HeLa S3 cells were pre-fractionated using custom two-dimensional liquid electrophoresis platform, comprising sIEF coupled to multiplexed GELFrEE[12,13]. HeLa S3, H1299, B16F10 cells and mitochondrial membrane proteins were also fractionated using the custom GELFrEE[13] device alone (no sIEF). After separation, detergent and salt were removed, and the fractions were injected into nanocapillary reversed-phase liquid chromatography (RPLC) columns for elution into a 12 T linear ion trap Fourier-transform mass spectrometer for online detection and fragmentation[14,15]. The mass spectrometry RAW files were processed with in-house software called CRAWLER to assign masses. Using this program, determination of both the intact masses and the corresponding fragment masses was performed and these data were searched against a human proteome database. Extensive statistical workups were also performed using several false discovery rate (FDR) estimation approaches (with decoy databases both concatenated and not). A final q value procedure is described in detail (Methods), with the data above reported using a 5% instantaneous FDR (that is, q value) cutoff at the protein level (Supplementary Fig. 14).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 March; accepted 15 September 2011.
Published online 30 October; corrected 7 December 2011 (see full-text HTML version for details).

1. Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nature Methods* **6,** 359–362 (2009).
2. Nesvizhskii, A. I. & Aebersold, R. Interpretation of shotgun proteomic data: the protein inference problem. *Mol. Cell. Proteomics* **4,** 1419–1440 (2005).
3. Schluter, H., Apweiler, R., Holzhutter, H. G. & Jungblut, P. R. Finding one's way in proteomics: a protein species nomenclature. *Chem. Cent. J.* **3,** 11 (2009).
4. Boyne, M. T., Pesavento, J. J., Mizzen, C. A. & Kelleher, N. L. Precise characterization of human histories in the H2A gene family by top down mass spectrometry. *J. Proteome Res.* **5,** 248–253 (2006).
5. Ge, Y., Rybakova, I. N., Xu, Q. G. & Moss, R. L. Top-down high-resolution mass spectrometry of cardiac myosin binding protein C revealed that truncation alters protein phosphorylation state. *Proc. Natl Acad. Sci. USA* **106,** 12658–12663 (2009).
6. Nilsson, T. *et al.* Mass spectrometry in high-throughput proteomics: ready for the big time. *Nature Methods* **7,** 681–685 (2010).
7. Mazur, M. T. *et al.* Quantitative analysis of intact apolipoproteins in human HDL by top-down differential mass spectrometry. *Proc. Natl Acad. Sci. USA* **107,** 7728–7733 (2010).
8. Righetti, P. G., Castagna, A., Antonioli, P. & Boschetti, E. Prefractionation techniques in proteome analysis: the mining tools of the third millennium. *Electrophoresis* **26,** 297–319 (2005).
9. Wang, H. *et al.* Intact-protein-based high-resolution three-dimensional quantitative analysis system for proteome profiling of biological fluids. *Mol. Cell. Proteomics* **4,** 618–625 (2005).
10. Capriotti, A. L., Cavaliere, C., Foglia, P., Samperi, R. & Laganà, A. Intact protein separation by chromatographic and/or electrophoretic techniques for top-down proteomics. *J. Chromatogr. A* (in the press).
11. Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y. & Aebersold, R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl Acad. Sci. USA* **97,** 9390–9395 (2000).
12. Tran, J. C. & Doucette, A. A. Multiplexed size separation of intact proteins in solution phase for mass spectrometry. *Anal. Chem.* **81,** 6201–6209 (2009).
13. Tran, J. C. & Doucette, A. A. Gel-eluted liquid fraction entrapment electrophoresis: an electrophoretic method for broad molecular weight range proteome separation. *Anal. Chem.* **80,** 1568–1573 (2008).

©2011 Macmillan Publishers Limited. All rights reserved

14. Lee, J. E. *et al.* A robust two-dimensional separation for top-down tandem mass spectrometry of the low-mass proteome. *J. Am. Soc. Mass Spectrom.* **20,** 2183–2191 (2009).
15. Vellaichamy, A. *et al.* Size-sorting combined with improved nanocapillary liquid chromatography-mass spectrometry for identification of intact proteins up to 80 kDa. *Anal. Chem.* **82,** 1234–1244 (2010).
16. Roth, M. J. *et al.* Precise and parallel characterization of coding polymorphisms, alternative splicing, and modifications in human proteins by mass spectrometry. *Mol. Cell. Proteomics* **4,** 1002–1008 (2005).
17. Durbin, K. R. *et al.* Intact mass detection, interpretation, and visualization to automate top down proteomics on a large scale. *Proteomics* **10,** 3589–3597 (2010).
18. Duncan, M. W., Aebersold, R. & Caprioli, R. M. The pros and cons of peptide-centric proteomics. *Nature Biotechnol.* **28,** 659–664 (2010).
19. Bunger, M. K., Cargile, B. J., Ngunjiri, A., Bundy, J. L. & Stephenson, J. L. J. Automated proteomics of *E. coli* via top-down electron-transfer dissociation mass spectrometry. *Anal. Chem.* **80,** 1459–1467 (2008).
20. Parks, B. A. *et al.* Top-down proteomics on a chromatographic time scale using linear ion trap fourier transform hybrid mass spectrometers. *Anal. Chem.* **79,** 7984–7991 (2007).
21. Patrie, S. M. *et al.* Top down mass spectrometry of 60-kDa proteins from *Methanosarcina acetivorans* using quadrupole FTMS with automated octopole collisionally activated dissociation. *Mol. Cell. Proteomics* **5,** 14–25 (2006).
22. Roth, M. J., Parks, B. A., Ferguson, J. T., Boyne, M. T. I. & Kelleher, N. L. 'Proteotyping': Population proteomics of human leukocytes using top down mass spectrometry. *Anal. Chem.* **80,** 2857–2866 (2008).
23. Gomez, S. M., Nishio, J. N., Faull, K. F. & Whitelegge, J. P. The chloroplast grana proteome defined by intact mass measurements from liquid chromatography mass spectrometry. *Mol. Cell. Proteomics* **1,** 46–59 (2002).
24. Matsuoka, S. *et al.* ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* **316,** 1160–1166 (2007).
25. Roberson, R. S., Kussick, S. J., Vallieres, E., Chen, S. Y. & Wu, D. Y. Escape from therapy-induced accelerated cellular senescence in p53-null lung cancer cells and in human lung cancers. *Cancer Res.* **65,** 2795–2803 (2005).
26. Yawata, T. *et al.* Identification of a $\leqq$ 600-kb region on human chromosome 1q42.3 inducing cellular senescence. *Oncogene* **22,** 281–290 (2003).
27. Sgarra, R. *et al.* During apoptosis of tumor cells HMGA1a protein undergoes methylation: identification of the modification site by mass spectrometry. *Biochemistry* **42,** 3575–3585 (2003).
28. Sgarra, R. *et al.* The AT-hook of the chromatin architectural transcription factor high mobility group A1a is arginine-methylated by protein arginine methyltransferase 6. *J. Biol. Chem.* **281,** 3764–3772 (2006).
29. Narita, M. *et al.* A novel role for high-mobility group A proteins in cellular senescence and heterochromatin formation. *Cell* **126,** 503–514 (2006).
30. Service, R. F. Proteomics ponders prime time. *Science* **321,** 1758–1761 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank all members of the group who contributed to development of top-down mass spectrometry over the years along with several private foundations: The Searle Scholars Program, The Burroughs Wellcome Fund, The David and Lucile Packard Foundation, The Richard and Camille Dreyfus Foundation, and The Chicago Biomedical Consortium with support from The Searle Funds at The Chicago Community Trust. We further acknowledge the Department of Chemistry at the University of Illinois, the Neuroproteomics Center on Cell to Cell Signaling supported through the National Institute on Drug Abuse (P30DA 018310), the National Institute for General Medical Sciences (GM 067193-08) and the National Science Foundation (DMS 0800631), whose combined investment in basic research over the past decade made this work possible. We dedicate this work in memory of Jonathan Widom.

Author Contributions Project design: J.C.T., L.Z., P.M.T., N.L.K. Cell culture and biology: J.C.T., J.E.L., A.D.C., D.R.A., M.L., C.W., S.M.M.S., N.S. Separations: J.C.T., J.E.L., A.D.C., D.R.A. Mass spectrometry: J.C.T., J.E.L., A.D.C., D.R.A., J.D.T., A.V., J.F.K., P.D.C. Data analysis and statistics: J.C.T., L.Z., K.R.D., B.P.E., R.D.L., P.M.T., N.L.K. Writing: J.C.T., N.L.K.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.L.K. (n-kelleher@northwestern.edu).

©2011 Macmillan Publishers Limited. All rights reserved

METHODS

Cell culturing and treatments. HeLa S3 cells (ATCC CCL-2.2) were grown in Joklik's modified minimal essential medium. HeLa cells were grown in suspension, whereas B16F10 mouse melanoma cells (ATCC CRL-6475) and H1299 small lung carcinoma cells (ATCC CRL-5803, in DMEM) were grown adherently. Media were supplemented with 10% newborn calf serum and 1% penicillin and streptomycin. Cells were maintained in 5% CO_2 at 37 °C, harvested (after adding 0.05% trypsin and EDTA for B16F10 and H1299 cells) by centrifugation at 200*g* for 5 min, and washed twice with PBS.

For experiments using HeLa with intrinsic DNA damage, cells were treated for 1 or 5 h using 1 μM etoposide (four-dimensional platform), or for 1 h with 25 or 100 μM etoposide for targeted experiments (three-dimensional platform, Supplementary Fig. 10 and Supplementary Table 6)[31]. Repair of DNA damage was monitored after placing treated (25 μM) cells into fresh media for 24 h before harvesting. B16F10 cells were treated with 10 μM etoposide for 5 h at 20–30% confluence, and allowed to grow in normal media. Similarly, H1299 cells were treated with 25 nM camptothecin for 24 h. Over several days, these treatments induced stress-associated, accelerated senescence as monitored by[32] a flattened and enlarged cell morphology, expression of senescence-associated β-galactosidase (SA-β-gal), formation of senescence-associated heterochromatic foci and upregulation of p53 (in B16F10 cells only). Approximately 2×10^7 senescent cells were collected, lysed and subjected to the three-dimensional platform.

Preparation of HeLa S3 cytosolic, nuclear and whole-cell extracts. For large-scale mapping using the four-dimensional platform, the HeLa cytosolic and nuclear extracts were prepared through a protocol by Trinkle-Mulcahy *et al.*[33]. After isolation, the pelleted fraction containing nuclei was redissolved using 4% SDS (50 mM Tris, pH 7.5, with protease, phosphatase inhibitors and sodium butyrate). Both cytosolic and nuclear fractions were disrupted using a sonication probe. The nuclear fractions were further homogenized using QIAShredder homogenizer spin columns to reduce viscosity (Qiagen). All fractions were centrifuged at 14,000*g* for 10 min at 4 °C. Protein concentrations were determined by bicinchoninic acid and stored at −80 °C.

For the three-dimensional platform, whole-cell extracts were re-suspended in 5 ml of lysis buffer (4% SDS, 100 mM Tris-HCl pH 7.5, 10 mM DTT with protease, phosphatase inhibitors and sodium butyrate). The mixture was vortexed for 5 min and boiled for 10 min. Immediately after boiling, the samples were alkylated in the dark with 100 mM iodoacetamide for 20 min.

Western blots, imaging and microscopy. Western Blots. Antibodies used were Histone H2A.X-pSer139 (Cell Signaling Technology; 2577S), GAPDH (Santa Cruz; sc-47724) and HRP-conjugated secondary antibodies. Chemiluminescence was detected using ChemiDoc XRS+ (Bio-Rad Laboratories) and band densities were calculated using Image Lab software (Bio-Rad Laboratories).

For γH2A.X imaging by immunofluorescence, cells were fixed in 0.1% glutaraldehyde and 3% formaldehyde made fresh from paraformaldehyde followed by permeabilization in 0.5% Triton X-100. Cells were subsequently incubated for 1 h at 25 °C with 3% BSA and then with the primary antibody against Histone H2A.X-pSer139 (1:400). Alexa Fluor Conjugates (Molecular Probes) were used as the secondary antibodies (1:1,000). Images were obtained using an inverted Zeiss Axio Observer.Z1 confocal microscope.

For β-Gal staining, cells were fixed for 5 min in 2% formaldehyde/0.2% glutaraldehyde in PBS, washed and incubated at 37 °C (no CO_2) overnight with X-Gal (5-bromo-4-chloro-3-indolyl β-D-glactosidase) staining solution (1 mg of X-Gal in 40 mM citric acid/sodium phosphate pH 6.0, 5 mM potassium ferrocyanide, 5 mM potassium ferricyanide, 150 mM NaCl, 2 mM $MgCl_2$).

For DAPI staining, cells were fixed in 4% paraformaldehyde for 30 min, washed and incubated with 4′,6′-diamidino-2-phenylindole (DAPI) solution (10 μg ml^{-1}) for 10 min. The images of DAPI-stained DNA were obtained using the Zeiss Axio Observer.Z1 confocal microscope.

Sample handling and multidimensional protein fractionation. Fractionation using four-dimensional fractionation (two-dimensional liquid electrophoresis and LC–MS). HeLa proteins (0.5–2 mg) were reduced, alkylated, precipitated with cold acetone and re-suspended in 3.2 ml sIEF buffer (8 M urea, 2 M thiourea, 50 mM DTT, 1% w/v Biolyte 3/10 carrier ampholytes from Bio-Rad Laboratories). The sample was focused using a custom designed eight-chamber sIEF system as previously described[34]. After complete focusing (approximately 1.5 h at 2 W), the liquid fractions (400 μl) were collected and combined with their respective chamber rinse solution (100 μl of 1% SDS). Adjacent sIEF fractions (including anode and cathode) were pooled, resulting in about five fractions that were precipitated using cold acetone. The precipitated proteins were re-suspended in approximately 50 μl of Laemmli loading buffer[35]. These fractions were then fractionated in parallel using a custom ten-channel multiplexed GELFrEE device[12]. Tube gels were cast to 12% T (1 cm length) for the resolving and 4% T for the stacking gels (300 μl volume). Application of 240 V for approximately 1 h resulted in eight or nine GELFrEE fractions (150 μl) per IEF fraction after elution of the dye front. After complete electroelution, the two-dimensional liquid electrophoresis fractions underwent SDS removal using chloroform/methanol/water precipitation as described previously[36]. Before nanocapillary RPLC injection, fractions were re-suspended by pipetting vigorously with 15–40 μl solvent A (5% acetonitrile, 0.2% formic acid).

For experiments using the three-dimensional platform, only GELFrEE coupled to nanocapillary LC–MS was used for sample fractionation. Whole cell lysates, mitochondrial membrane preparations or extracts targeting modified proteins were re-suspended as described above and fractionated using a single-channel GELFrEE device[13]. Nanocapillary LC–MS conditions for large-scale analyses were as described below.

For nanocapillary RPLC–MS, in either the four- or three-dimensional platform, re-suspended fractions were injected (10 μl) onto a trap column (150 μm × 2 cm) using an autosampler (Eksigent). The nanobore analytical column (75 μm × 10 cm) containing an integral fritted nanospray emitter (PicoFrit, New Objective) was coupled to the trap in a vented column tee setup. Both the analytical and trap columns contain polymeric reversed-phase (PLRP-S, Phenomenex) media (5 μm, 1,000 Å pore size). The Eksigent 1D Plus nano-HPLC system was operated at a flow rate of approximately 2 μl min^{-1} for 10 min for loading onto the trap. The proteins were eluted into the mass spectrometer using a flow rate of 300 nl min^{-1} with the following gradient: 5% B (95% acetonitrile + 0.2% formic acid) at 0 min; 20% B at 5 min; 55% B at 50 min; 85% B at 55 min; 5% B at 65 min; 5% B at 75 min.

For proteins fractionated from HeLa S3 and B16F10 cells, the nanocapillary RPLC column was coupled online to a 12 T LTQ FT Ultra mass spectrometer (Thermo Fisher Scientific) fitted with a digitally controlled nanospray ionization source (PicoView DPV-550, New Objective). For masses up to 25 kDa (as determined from GELFrEE fractions), precursor mass data were collected using the Fourier-transform ion cyclotron resonance (eight microscans, 170,000 resolving power at $m/z = 400$) with an m/z range of 500–1,800 and a target value of 1 million charges. For masses either greater than 25 kDa or greater than 50 kDa, precursor mass data were obtained using the ion trap at 20 or 50 microscans, respectively. Data from H1299 cells were obtained on an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific) using two to four microscans, 108,000 or 216,000 resolving power at $m/z = 400$, and a target value of 1 million charges.

Mass spectrometry data acquisition for targeted monitoring of intact isoforms/species. For three-dimensional experiments on targets up to 25 kDa, precursor mass data were collected with Fourier-transform ion cyclotron resonance parameters as described above. Data-dependent 'zoom mapping' was performed using a top three (no fragmentation) acquisition strategy with 60 m/z isolation window, three microscans at 85,000 resolving power (target value of 2 million charges with 60 m/z isolation at zero collision energy, or SIM mode). Dynamic exclusion was enabled with a repeat count of 2, an exclusion duration of 5,000 s, and a repeat duration of 240 s. 'Mass mode' was enabled in the Xcalibur software to ensure that each zoom map scan detected a different protein species.

Preparation of mitochondrial membrane proteins for top-down mass spectrometry. A HeLa S3 cell pellet consisting of approximately 10^9 cells was re-suspended in about 16 ml STM buffer (250 mM sucrose, 50 mM Tris-HCl, pH 7.4, 5 mM $MgCl_2$, 10 mM sodium butyrate, 1 mM DTT, 1% protease and phosphatase inhibitors (Sigma-Aldrich)). Cells were lysed using a glass Dounce homogenizer. The lysate was centrifuged at 800*g* for 15 min to remove nuclei and cellular debris. Mitochondria membrane isolation was performed as described previously[37]. Briefly, the cell lysate was centrifuged at 6,000*g* for 15 min. The pellet was washed with STM buffer and the centrifugation step repeated. Mitochondria were re-suspended in 2 ml lysis buffer (10 mM HEPES, pH 7.9, 10 mM sodium butyrate, 1 mM DTT, 1% protease and phosphatase inhibitors) and stirred at 4 °C for 30 min before sonication. The suspension was centrifuged at 9,000*g* for 30 min. The pellet was re-suspended in 0.5 ml extraction buffer (20 mM Tris, pH 7.8, 0.4 M NaCl, 15% glycerol, 5% SDS, 10 mM sodium butyrate, 1 mM DTT, 1% protease and phosphatase inhibitors) and vortexed for 30 min. The sample was centrifuged at 9,000*g* for 30 min and the supernatant was collected, aliquoted, flash frozen and stored at −80 °C until use.

For GELFrEE separation and LC–MS of integral membrane proteins, mitochondrial membrane proteins (400 μg as determined by bicinchoninic acid) were acetone precipitated, re-suspended in 100 μl loading buffer, reduced with 20 mM DTT and alkylated with 100 mM iodoacetamide. After a GELFrEE separation as reported previously[13], SDS was removed[36] and the fractions re-suspended in 30 μl of fresh 60% formic acid. Each 10-μl fraction was injected onto a PLRP analytical column (75 μm × 10 cm) and heated to 45 °C using a nanocapillary column heater (New Objective). A solvent system developed for integral membrane proteins was used in nanocapillary format here: A, 60% formic acid in water; B, 100% isopropanol. The gradient used was 0% B at 0 min, 25% B at 5 min, 60% B at 50 min, 95% B at 56 min. Ten GELFrEE samples were also analysed by the standard acetonitrile solvent and gradient system described above.

©2011 Macmillan Publishers Limited. All rights reserved

TMHMM version 2.0 (http://www.cbs.dtu.dk/services/TMHMM) was used for the transmembrane domain prediction of identified proteins. The SwissProt accession number for each identification was entered into the server to determine the number and location of all transmembrane domains. Differences between the database sequence and that identified by top-down mass spectrometry were accounted for by interrogating the identified sequence with TMHMM.

Protein identification and characterization. Mass spectrometry fragmentation data were acquired in three different modes, depending on protein mass range. Below 17 kDa, data-dependent collision-induced dissociation (CID) was used (FT/FT), whereas source-induced dissociation (SID) was used for masses between 17 and 25 kDa (FT/FT) and above 25 kDa (ion trap/FT). Based on preliminary analyses, SID of 15 V was optimal for ion trap scans for the dissociation of weakly bound non-covalent adducts, whereas the 75 V SID for fragmentation was standardized as described[15]. For data-dependent fragmentation (top two precursors, 15–25 m/z isolation window), dynamic exclusion was enabled with a repeat count of 2, an exclusion duration of 5,000 s, and a repeat duration of 240 s. Both CID and SID were collected using eight microscans and 85,000 resolving power (at $m/z = 400$) with a target value of 2 million charges. Data for H1299 cells obtained on the Orbitrap Elite were acquired using either higher-energy collisional dissociation, CID or electron transfer dissociation during top three data-dependent tandem mass spectrometry fragmentation using mass mode, four microscans, 108,000 resolving power and a target value of 1 million charges (cf. Fig. 4i, j, Supplementary Figs 12 and 13 and Supplementary Table 7).

Software and data analysis. Much of the software for intact mass determination (kDECON), generation of visual outputs (Proteome Display), and species differentiation (PTMcrawler) has been published recently[17]. Briefly, kDECON provides average mass information for charge state resolved distributions using the ion trap. PTMcrawler traverses a list of intact monoisotopic masses to find mass differences corresponding to PTMs such as methylations, acetylations and phosphorylations (Fig. 2b). Proteome Display was devised for visualization of nanocapillary LC–tandem mass spectrometry data, allowing graphical viewing of four-dimensional proteome runs and generation of plots for specific PTMs (Fig. 2).

The RAW files collected were first processed with an algorithm called CRAWLER to assign masses. Using a version of this program, scans collected by Fourier-transform mass spectrometry were de-isotoped using the Xtract or THRASH algorithm[38]. Scans collected in the ion trap were deconvoluted using the kDECON algorithm (minimum intensity cutoff 1,000; mass range 10–70 kDa). Xtract or THRASH processing generated monoisotopic neutral masses, whereas kDECON processing provided average neutral masses. Data-dependent fragmentation scans were summed within a retention time tolerance of 1 min and a precursor tolerance of 0.05 m/z whereas SID scans were summed in 0.3 min. In both cases, multiplexed fragmentation was considered. Fragmentation data were filtered by selecting the top three most intense neutral fragment masses within a 100-Da window below 2,000 Da, and the top five above 2,000 Da. One or more precursor masses and one or more fragment masses resulting from each summed unit were grouped as ProSightPC experiment, and written to a ProSight Upload Format file. If a precursor mass could not be determined, the experiment was written out to a separate file with a placeholder value, for separate analysis.

The ProSight Upload Format files output by CRAWLER were searched against a human proteome database using a custom implementation of ProSightPC 2.0 with iterative search logic[39–41] on a 168-node Rocks[42] cluster. Four types of analysis were run, depending on the type of data: FT/FT CID, FT/FT SID, ion trap/FT SID or SID data where the precursor could not be determined ('No-Hi-SID'). The iterative search trees were designed to take advantage of high mass accuracy, while retaining the option to run less specific searches if a result of sufficient quality could not be obtained by more specific searches. All searches used 10 p.p.m. tolerance for the fragment ions, all of which were obtained at high resolving power. For each search, the top ten hits were returned; if the top hit had an E value not more than 1×10^{-2}, the analysis moved on to the next experiment, otherwise the next search in the tree was run. All searches were in absolute mass mode. The FT–FT–CID tree consisted of searches at 200, 2,000 Da, and 'entire database' precursor tolerances; the FT–FT–SID at 2.3, 2,000, 20,000 and entire database; and ion trap–FT–SID trees used 2,000, 20,000 Da, and 'entire database' precursor tolerances; the No-Hi-SID tree just searched against the entire database. Searches were against two different human proteome databases built against UniProt release 2011_04, encompassing known alternative splices, modifications, peptide cleavage events, potential initial methionine cleavage and amino (N)-terminal acetylation. A complex database was created encompassing combinations of annotated alternative splice and peptide cleavage events to generate 54,190 base sequences. A maximum of 2^{13} protein forms for each base sequence make approximately 8,450,000 theoretical protein species. This database was used for all searches where the precursor tolerance was less than 2,000 Da. All other searches used a simplified database consisting of the same 54,190 base sequences, modified with N-terminal acetylation and initial methionine cleavage (where applicable) creating a total of approximately 160,000 forms.

Data were run against both forward and scrambled databases (see FDR estimation below) separately with identical search parameters. Upon completion, all search results were loaded into a ProSight data repository and a report was produced, returning the top hits for each experiment. The hit with the best q value was then chosen as the exemplar for each gene product cluster. If two member hits had the same q value, the member with the lowest absolute mass difference to the theoretical hit was chosen, and if this still produced a set with more than one member, the form with the 'most-terminal' PTMs (that is, closest to the N or carboxy terminus) was chosen. For a desired FDR cutoff (for example, 5%), a list of accession numbers and species is produced.

For estimation of FDR for top-down proteomics, the Poisson-based model as published in 2001 (ref. 43) had been previously modified with a Bonferroni correction that enables probability-based scoring (that is, use of the E value noted above) for searches done on a database created by shotgun annotation[16]. To validate the process used in high-throughput operations, an FDR analysis was performed to correct for multiple hypothesis testing using the method of Benjamini and Hochberg[44] as applied by Storey[45]. For each Poisson-based p value, a corresponding Bayesian posterior p value is calculated, termed the q value, which is a measure of the FDR for that particular identification event (also called an instantaneous FDR).

To calculate q values, a separate decoy database of scrambled sequences was created equal in size to the forward (real) database[46]. Searches were done separately on both the forward and decoy (scrambled) databases using all data in a set of four- or three-dimensional proteome runs. A histogram was created for the decoy database results using the log p_{id} value (where p_{id} is the Poisson probability of an incorrect protein identification) of these false identifications (Supplementary Fig. 14). These data were modelled against a gamma distribution and fit with a shape (k) of 10.26 ± 0.04 and rate (θ) of 2.25 ± 0.01. The distribution of scrambled hits is taken as an empirical estimate of the distribution of scores under the null hypothesis that the match was due to chance. Thus, the area under the scrambled score distribution to the right of the observed forward score is the probability of getting this good a forward score, or better, by chance (abbreviated hereafter p_F). From here, all data are rank-ordered by their corresponding p_F values and q values are calculated as in Storey[45]. The final results were generated using a q value cutoff of 0.05, thus achieving a protein level FDR of 5%. For comparison, bottom-up studies typically use a 1% FDR cutoff at the peptide level (the so-called PSM level, for peptide spectrum match), which typically rolls up into a 5–8% FDR at the protein level[47]. Extensive comparison with other FDR estimation techniques (including generating ROC curves for reversed decoy databases of concatenated reversed sequences) showed the q value approach to be 10–25% more stringent in terms of number of identified proteins and species.

31. Soubeyrand, S., Pope, L. & Hache, R. J. G. Topoisomerase II α-dependent induction of a persistent DNA damage response in response to transient etoposide exposure. *Mol. Oncol.* **4,** 38–51 (2010).
32. Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16(INK4a). *Cell* **88,** 593–602 (1997).
33. Trinkle-Mulcahy, L. *et al.* Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. *J. Cell Biol.* **183,** 223–239 (2008).
34. Tran, J. C. & Doucette, A. A. Rapid and effective focusing in a carrier ampholyte solution isoelectric focusing system: a proteome prefractionation tool. *J. Proteome Res.* **7,** 1761–1766 (2008).
35. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of Bacteriopage T4. *Nature* **227,** 680–685 (1970).
36. Wessel, D. & Flugge, U. I. A Method for the Quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* **138,** 141–143 (1984).
37. Cox, B. & Emili, A. Tissue subcellular fractionation and protein extraction for use in mass-spectrometry-based proteomics. *Nature Protocols* **1,** 1872–1878 (2006).
38. Horn, D. M., Zubarev, R. A. & McLafferty, F. W. Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J. Am. Soc. Mass Spectrom.* **11,** 320–332 (2000).
39. Boyne, M. T. *et al.* Tandem mass spectrometry with ultrahigh mass accuracy clarifies peptide identification by database retrieval. *J. Proteome Res.* **8,** 374–379 (2009).
40. LeDuc, R. D. *et al.* ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry. *Nucleic Acids Res.* **32,** W340–W345 (2004).
41. Zamdborg, L. *et al.* ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res.* **35,** W701–W706 (2007).
42. Papadopoulos, P. M., Katz, M. J. & Bruno, G. in *Proc. 3rd IEEE Int. Conf. Cluster Computing* 258 ⟨http://www.computer.org/portal/web/csdl/doi/10.1109/CLUSTR.2001.959986⟩ (2001).

©2011 Macmillan Publishers Limited. All rights reserved

43. Meng, F. Y. *et al.* Informatics and multiplexing of intact protein identification in bacteria and the archaea. *Nature Biotechnol.* **19,** 952–957 (2001).
44. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate – a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57,** 289–300 (1995).
45. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100,** 9440–9445 (2003).
46. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods* **4,** 207–214 (2007).
47. Reiter, L. *et al.* Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry. *Mol. Cell. Proteomics* **8,** 2405–2417 (2009).

©2011 **Macmillan Publishers Limited. All rights reserved**

LETTER

doi:10.1038/nature10547

The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements

Serena De Fazio[1], Nenad Bartonicek[2]*, Monica Di Giacomo[1]*, Cei Abreu-Goodger[2], Aditya Sankar[1], Charlotta Funaya[3], Claude Antony[3], Pedro N. Moreira[1], Anton J. Enright[2] & Dónal O'Carroll[1]

Piwi proteins and Piwi-interacting RNAs (piRNAs) have conserved functions in transposon silencing[1]. The murine Piwi proteins Mili and Miwi2 (also called Piwil2 and Piwil4, respectively) direct epigenetic LINE1 and intracisternal A particle transposon silencing during genome reprogramming in the embryonic male germ line[2–4]. Piwi proteins are proposed to be piRNA-guided endonucleases that initiate secondary piRNA biogenesis[5–7]; however, the actual contribution of their endonuclease activities to piRNA biogenesis and transposon silencing remain unknown. To investigate the role of Piwi-catalysed endonucleolytic activity, we engineered point mutations in mice that substitute the second aspartic acid to an alanine in the DDH catalytic triad of Mili and Miwi2, generating the *Mili*^{DAH} and *Miwi2*^{DAH} alleles, respectively. Analysis of Mili-bound piRNAs from homozygous *Mili*^{DAH} fetal gonadocytes revealed a failure of transposon piRNA amplification, resulting in the marked reduction of piRNA bound within Miwi2 ribonuclear particles. We find that Mili-mediated piRNA amplification is selectively required for LINE1, but not intracisternal A particle, silencing. The defective piRNA pathway in *Mili*^{DAH} mice results in spermatogenic failure and sterility. Surprisingly, homozygous *Miwi2*^{DAH} mice are fertile, transposon silencing is established normally and no defects in secondary piRNA biogenesis are observed. In addition, the hallmarks of piRNA amplification are observed in Miwi2-deficient gonadocytes. We conclude that cycles of intra-Mili secondary piRNA biogenesis fuel piRNA amplification that is absolutely required for LINE1 silencing.

Transposable elements are mobile genetic elements that constitute a large fraction of eukaryotic genomes. The process of transposon silencing is of fundamental importance for genome integrity and germ cell development. Members of the Piwi subclade of the Argonaute proteins have conserved roles in transposon silencing and bind a class of small non-coding RNAs known as piRNAs that act as guides for targeting of the respective ribonuclear particles (RNPs)[1]. The Argonaute family is primarily defined by the presence of the Piwi domain that adopts a classical RNase H fold[8] with some Argonaute proteins being active small-RNA-guided endonucleases (slicers)[9]. Mechanisms of piRNA biogenesis are largely unclear, but two models summarize our current knowledge. First, primary processing of long single-stranded precursors by unknown nuclease(s) results in the generation of primary piRNAs. Thereafter, secondary biogenesis takes place via a 'ping-pong' cycle; that is, a feed-forward amplification loop wherein the slicer activities of two Piwi proteins take turns to generate secondary piRNAs using primary piRNAs as initial guides[5,6]. Originally described in *Drosophila*, where all three *Drosophila* Piwi proteins have proven endonuclease activity *in vitro*[6,10], this secondary pathway model offers an explanation for transposon silencing as piRNA biogenesis consumes transposon transcripts[5,6]. Nevertheless, the role of Piwi-mediated slicing in secondary biogenesis or transposon silencing is not directly tested in any system.

In mice, transposable elements are epigenetically silenced throughout most of life with transposon repression being initially established during germ cell development[11]. The process of reprogramming in the fetal germ line initiates genome-wide CpG demethylation that erases both genomic imprints as well as retrotransposon promoter methylation[11]. In male mice the Piwi proteins Mili and Miwi2 are essential for *de novo* DNA methylation and repression of (L1) and intracisternal A particle (IAP) retrotransposons[2–4]. The mouse piRNA model proposes that transposon-derived primary Mili-bound sense piRNAs initiate a round of secondary processing by targeting complementary antisense transposon transcripts for cleavage by Mili's endonuclease activity[7]. The 3′ cleavage fragment is then 3′ processed, which results in the accumulation of secondary antisense piRNAs in Miwi2 that guide transposon DNA methylation. Furthermore, these

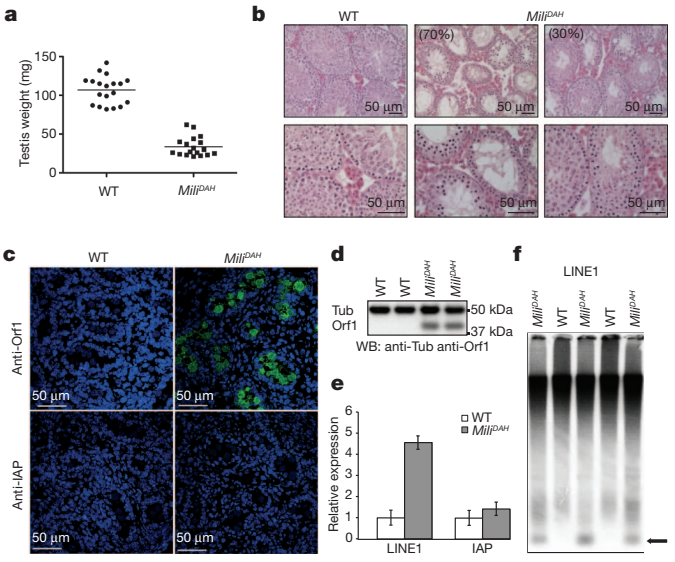


Figure 1 | The endonuclease activity of Mili is required for spermatogenesis and L1 silencing. **a**, Testicular atrophy in *Mili*^{DAH} mice. Testicular weights of 3-month-old wild-type (WT) and *Mili*^{DAH} mice are shown. **b**, Haematoxylin- and eosin-stained wild-type and *Mili*^{DAH} testis sections from 3-month-old mice. The percentage of *Mili*^{DAH} mice with the indicated phenotype is shown. **c**, Immunofluorescence using anti-L1 Orf1 and anti-IAP Gag antibodies (green) and DAPI-stained DNA (blue) on wild-type and *Mili*^{DAH} E16.5 fetal testis sections are shown. **d**, Western blot (WB) using anti-L1 Orf1 and anti-tubulin antibodies on extracts from P10 wild-type and *Mili*^{DAH} testes is shown. **e**, The expression levels of L1 and IAP were quantified by qRT–PCR from RNA derived from P10 wild-type and *Mili*^{DAH} testes. Error bars indicate standard deviation from biological triplicates ($n = 3$). **f**, Methylation-sensitive Southern blot on HpaII-digested DNA extracted from P10 wild-type and *Mili*^{DAH} testis using a L1 promoter probe. The arrowhead indicates the identity of the methylation-sensitive fragment.

[1]European Molecular Biology Laboratory, Mouse Biology Unit, Via Ramarini 32, Monterotondo Scalo 00015, Italy. [2]European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge CB10 1SD, UK. [3]European Molecular Biology Laboratory, EMBL Meyerhof Str. 1, 69117 Heidelberg, Germany.
*These authors contributed equally to this work.

©2011 Macmillan Publishers Limited. All rights reserved

antisense piRNA-programmed Miwi2 RNPs are also proposed to target transposon transcripts for cleavage followed by another round of secondary piRNA biogenesis that serves to amplify the primary transposon piRNAs[7]. Here we explore the contribution of the putative endonuclease activities of Mili and Miwi2 to the embryonic piRNA pathway and transposon silencing.

We decided to use the same genetic strategy that has recently been used to explore the *in vivo* endonuclease function of Argonaute 2 (Ago2)[12–14]. Mutation of the second aspartic acid to an alanine in the catalytic triad DDH (the $Ago2^{D669A}$ or $Ago2^{DAH}$ mutation/allele) abrogates Ago2's catalytic function without affecting protein stability and other functions of the protein *in vivo*[13–16]. We first generated knock-in $Mili^{D813A/D813A}$ or $Mili^{DAH}$ mice (Supplementary Fig. 1a–c). The point mutation did not have an impact on Mili protein expression (Supplementary Fig. 1d). Mili localizes to the inter-mitochondrial cement, a part of the germ-cell-specific perinuclear nuage[17], and has been shown to interact with the Tudor domain protein Tdrd1 as well as the RNA helicases Mov10L1 and Mvh (refs 18–22). Co-localization of Mili with Tdrd1, Mov10L1 and Mvh was observed in fetal gonadocytes of wild-type and $Mili^{DAH}$ mice (Supplementary Fig. 2). Furthermore, interaction of Mili with Tdrd1 and Mov10L1 was confirmed in both wild-type and $Mili^{DAH}$ mice (Supplementary Table 1). In contrast to Mili, Tdrd1 or Mvh deficiency[23,24], normal inter-mitochondrial cement was observed in $Mili^{DAH}$ gonadocytes (Supplementary Fig. 3). Therefore, the $Mili^{DAH}$ point mutation does not have an impact on inter-mitochondrial cement integrity, expression levels or localization of Mili.

Deletion of Mili or Miwi2 leads to arrest in meiotic prophase[3,25], probably attributable to secondary consequences of defective transposon DNA methylation and de-repression. All $Mili^{DAH}$ testes were atrophic (Fig. 1a), with most (70%) of $Mili^{DAH}$ mice presenting the same meiotic phenotype as $Mili^{-/-}$ mice. Early pachytene was the most advanced stage of germ cell development observed in these mice (Fig. 1b and Supplementary Fig. 4). The other 30% of $Mili^{DAH}$ mice presented a slightly milder phenotype. Tubules containing round and even elongating spermatids were observed, albeit with cells of aberrant morphology (Fig. 1b and Supplementary Fig. 5). All $Mili^{DAH}$ tubules were highly apoptotic (Supplementary Figs 4b and 5c). However, independent of the phenotype observed, all $Mili^{DAH}$ male mice are sterile (Supplementary Fig. 1e). The process of transposon *de novo* DNA methylation occurs during late gestation in mice, concluding in postnatal germ cells a few days after birth. De-repression of L1 elements was observed in $Mili^{DAH}$ embryonic day (E)16.5 gonadocytes and postnatal and adult germ cells (Fig. 1c–e and Supplementary Fig. 6a, b). With respect to the process of transposon silencing, no heterogeneity in adult $Mili^{DAH}$ mice was observed (Supplementary Fig. 6), raising the possibility that the distinct adult spermatogenic phenotypes may arise from a genetic background effect or that Mili and its associated endonuclease activity may function in processes beyond transposon silencing in meiotic or post-meiotic cells. Notably, in contrast to $Mili^{-/-}$

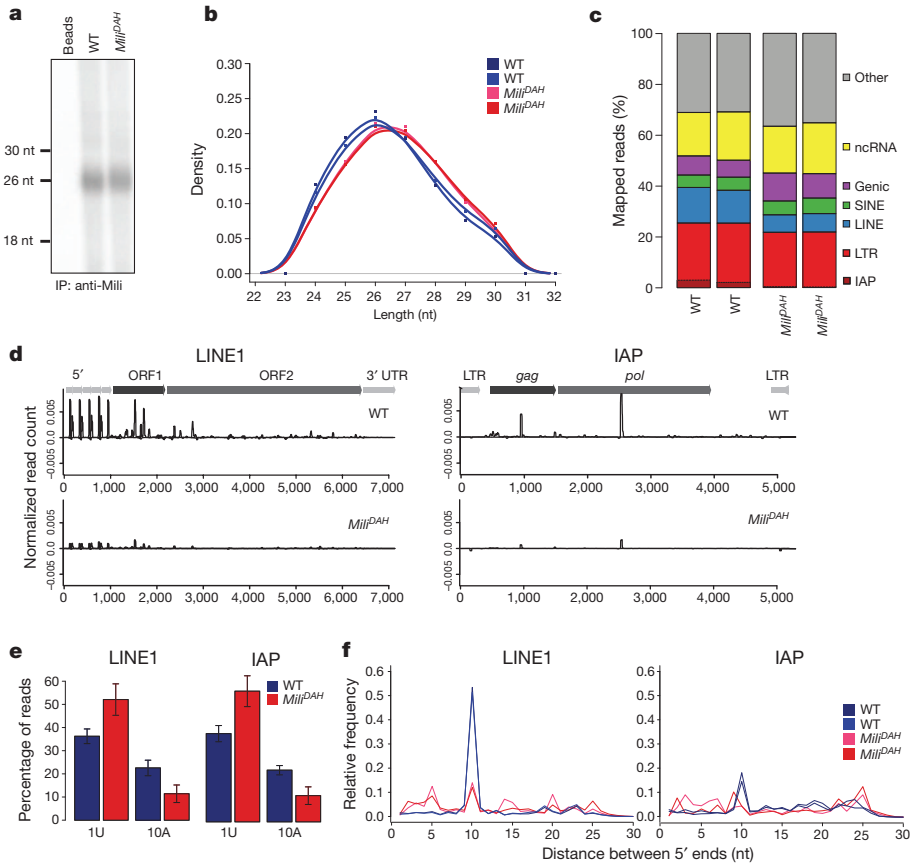


Figure 2 | piRNA amplification failure in $Mili^{DAH}$ mice. **a**, Mili RNPs were immunoprecipitated from E16.5 fetal testis of the indicated genotypes. Associated piRNAs were visualized by 5′ ^{32}P labelling, resolution on a 15% TBE urea gel and autoradiography. nt, nucleotide. **b**, Size profiles of cloned Mili-bound piRNAs from biological replicates of wild-type (blue) and $Mili^{DAH}$ (pink) E16.5 fetal gonadocytes. **c**, Genomic annotation of cloned Mili-bound piRNAs as indicated from pairs of biological replicates of wild-type and $Mili^{DAH}$ E16.5 fetal gonadocytes. LTR, long terminal repeat; ncRNA, non-coding RNA. **d**, Mapping of Mili-bound piRNAs to the consensus of L1 (left) and IAP (right) elements. Positive and negative values indicate sense and antisense piRNAs, respectively. Schematic representations of the respective elements are also shown (above). **e**, Percentage of piRNAs from wild-type and $Mili^{DAH}$ Mili RNPs with a U at the first position (1U) without A at position 10 and an A at position 10 (10A) without a U at position 1 are shown for L1 and IAP elements. Error bars indicate standard error of the mean from the biological duplicates ($n = 2$). **f**, Ping-pong analysis of Mili-bound piRNAs from biological replicates of wild-type and $Mili^{DAH}$ E16.5 fetal gonadocytes. Relative frequency (y axis) of distances between 5′ ends (x axis) between complementary piRNAs for both L1 and IAP elements is shown. nt, nucleotides.

©2011 Macmillan Publishers Limited. All rights reserved

mice[2,4], IAPs are normally repressed in *Mili*^{DAH} mice (Fig. 1c–e and Supplementary Fig. 6a, b). Accordingly, defective CpG DNA methylation of L1 but not IAP elements was observed in postnatal germ cells at time points after the completion of the *de novo* DNA methylation process (Fig. 1f and Supplementary Fig. 6c). Collectively, these data indicate that the endonuclease activity of Mili is absolutely required for L1 silencing.

We next analysed Mili-bound piRNAs to understand the impact of the *Mili*^{DAH} mutation on the piRNA pathway. Equal quantities of piRNA were observed in Mili RNPs from *Mili*^{DAH} fetal gonadocytes (Fig. 2a). piRNA libraries from biological duplicates of Mili RNPs from wild-type and *Mili*^{DAH} fetal gonadocytes were subjected to deep sequencing. 20.3% of piRNAs are longer in *Mili*^{DAH} versus wild-type gonadocytes (Fig. 2b). Among repetitive element piRNAs a 4.2- and 5.1-fold reduction in the fraction of L1 and IAP piRNAs, respectively, are observed in *Mili*^{DAH} libraries (Fig. 2c). When we mapped the reads to the genome, no differences were observed in the genomic origin between wild-type and *Mili*^{DAH} piRNAs (Supplementary Fig. 7), indicating that the endonuclease activity of Mili is not required for the selective expression of specific piRNA clusters. We next performed a qualitative analysis of L1 and IAP piRNAs. For this analysis, we only consider unique reads and not their individual depths within libraries. No differences in L1 and IAP piRNA formation were observed in the wild-type and *Mili*^{DAH} libraries (Supplementary Fig. 8a). We therefore conclude that the *Mili*^{DAH} mutation does not affect primary biogenesis of transposon or cluster-related piRNAs.

The precise endonucleolytic cleavage site of an Argonaute protein is at the nucleotide opposed to the tenth nucleotide from the 5′ end of the guide small-RNA[26,27]. A 5′U (1U) bias is a characteristic of primary piRNAs, therefore when a primary piRNA directs cleavage of a target followed by 3′ processing, the secondary piRNA will contain a bias for A at position 10 (refs 5, 6). In addition, there is an overlap of 10 nucleotides; this precise complementarity is detected when comparing distances between 5′ ends of complementary piRNAs and is known as the ping-pong signature[5,6]. A second round of piRNA biogenesis fuelled by the secondary piRNA results in the amplification of the initiating primary piRNA. Analysis of Mili-bound piRNAs, now taking depth into consideration, revealed reduced levels of both L1 and IAP piRNAs in *Mili*^{DAH} fetal gonadocytes (Fig. 2c, d). The increase in 1U and a decrease in 10A-containing piRNAs in *Mili*^{DAH} libraries are indicative of the failure of piRNA amplification (Fig. 2e). Notably, the ping-pong signature is lost for both L1 and IAP piRNAs in *Mili*^{DAH} mice (Fig. 2f). Collectively, these data are consistent with a function of the endonuclease activity of Mili in the initiation of secondary processing required for transposon piRNA amplification.

The lack of secondary processing in *Mili*^{DAH} gonadocytes resulted in the stark reduction of piRNA in Miwi2 RNPs (Fig. 3a). The additional cycles required to generate libraries from *Mili*^{DAH} gonadocytes supported the observation of a significant decrease in piRNAs within Miwi2 from *Mili*^{DAH} gonadocytes (Supplementary Table 2). Miwi2 localizes to cytoplasmic P-bodies known as piP bodies and to the nucleus[7,17]. Staining with the piP-body components Tdrd9 and

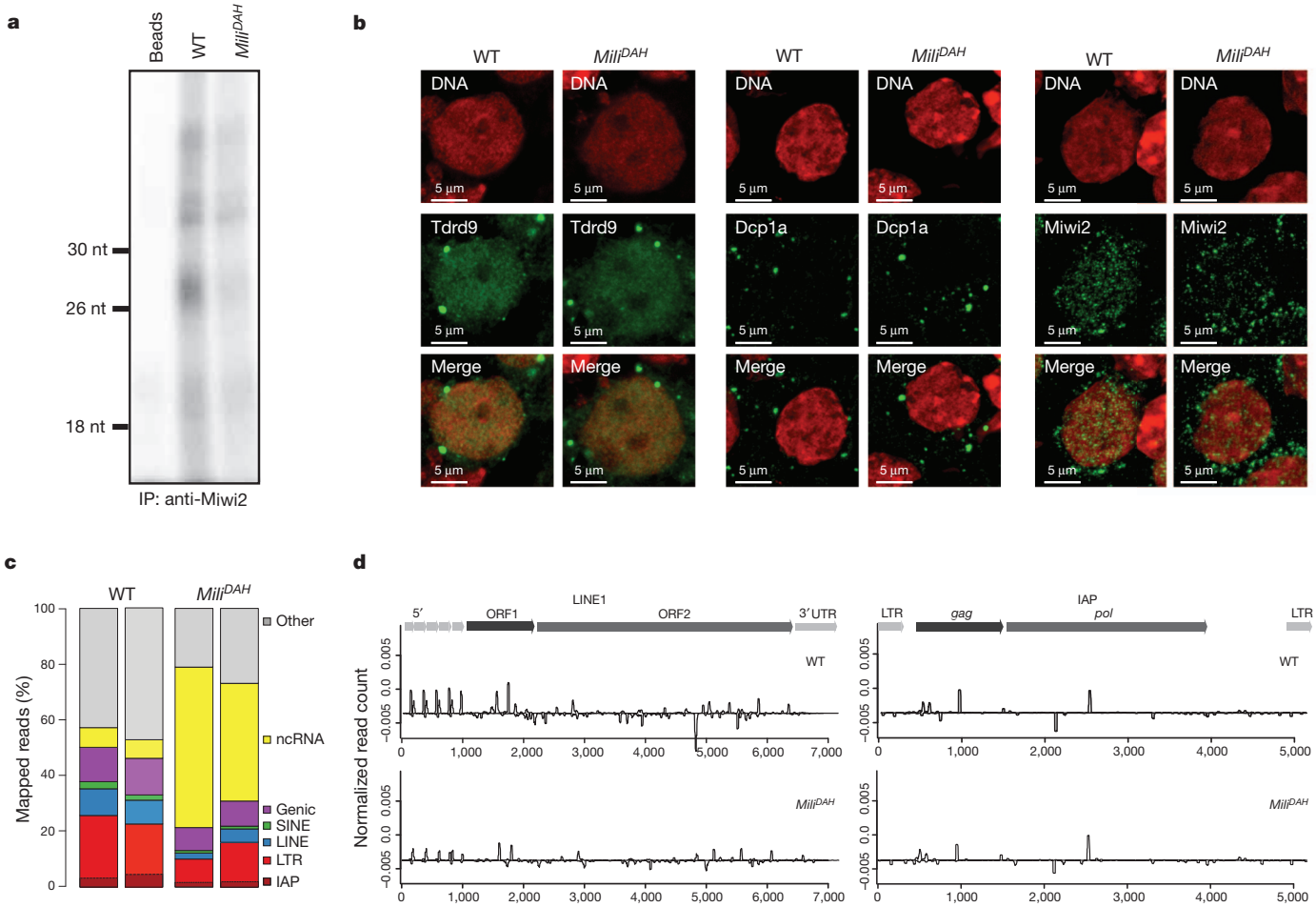


Figure 3 | Marked reduction of Miwi2-bound piRNAs in *Mili*^{DAH} mice. **a**, Miwi2 RNPs were immunoprecipitated from E16.5 fetal testis of the indicated genotypes and piRNA represented as in Fig. 2a. **b**, Confocal projection images of indirect immunofluorescence with Tdrd9, Dcp1a and Miwi2 antibodies (green) and DAPI-stained DNA (red) from E16.5 fetal testis of the indicated genotypes. **c**, Annotation of cloned Miwi2-bound piRNAs as indicated from biological replicates of wild-type and *Mili*^{DAH} E16.5 fetal gonadocytes. **d**, Mapping of the Miwi2-bound piRNAs to the consensus of L1 (left) and IAP (right) elements. Positive and negative values indicate sense and antisense piRNAs, respectively. Schematic representations of the respective elements are also shown.

©2011 Macmillan Publishers Limited. All rights reserved

Dcp1a revealed intact piP-body formation in *Mili*^{DAH} gonadocytes (Fig. 3b). Furthermore, Miwi2 shows cytoplasmic foci and, consistent with the partial piRNA loading, retained nuclear localization in *Mili*^{DAH} gonadocytes albeit with a lower staining intensity (Fig. 3b). Sequencing revealed a piRNA population of a normal size profile featuring an approximately 2-fold decrease in the percentage of both LINE- and IAP-associated piRNAs (Fig. 3c and Supplementary Fig. 9a). However, when mapped to L1 and IAP consensus sequences, the Miwi2-bound piRNAs from *Mili*^{DAH} mice showed a pattern of piRNA formation equivalent to wild type (Fig. 3d and Supplementary Fig. 8b). Whereas Miwi2 bound sense and antisense transposon piRNAs in almost equal measure in both wild-type and *Mili*^{DAH} gonadocytes (Supplementary Table 2), as previously described[4,7], the amplitude of the antisense piRNA peaks derived across L1 was diminished compared to those previously observed[4,7]. This difference probably reflects experimental variation due to the antibody used or recent advances in library generation/sequencing. No difference in the genomic origin of cluster-related piRNAs was observed (Supplementary Fig. 9b). Therefore, the endonuclease activity of Mili does not alter the identity of piRNAs that reside within Miwi2 RNPs, merely their quantity.

The fact that Miwi2 is seeded with the appropriate but severely reduced quantities of piRNAs in *Mili*^{DAH} mice indicates that the potential endonuclease activity of Miwi2 cannot compensate for the loss of Mili's slicer activity. To understand the contribution of Miwi2's putative endonuclease activity to piRNA amplification and transposon silencing, we generated *Miwi2*$^{D718A/D718A}$ (*Miwi2*^{DAH}) and *Miwi2*$^{-/-}$ mice (Supplementary Figs 10 and 11). Surprisingly, *Miwi2*^{DAH} mice are fertile with no defects observed in testis morphology (Fig. 4a). Furthermore, in contrast to *Miwi2*$^{-/-}$ fetal gonadocytes[3,4], both L1 and IAP are normally repressed (Fig. 4b, c). In addition to transposon silencing, Miwi2 is specifically required for the maintenance of spermatogonial stem cells[3]. No loss of germ cells is observed in *Miwi2*^{DAH} aged mice (9–12 months) (Supplementary Fig. 12). Therefore, the *Miwi2*^{DAH} mutation does not have an impact on any known physiological function of Miwi2. Accordingly, both Mili and Miwi2 are normally loaded with piRNA in *Miwi2*^{DAH} fetal gonadocytes

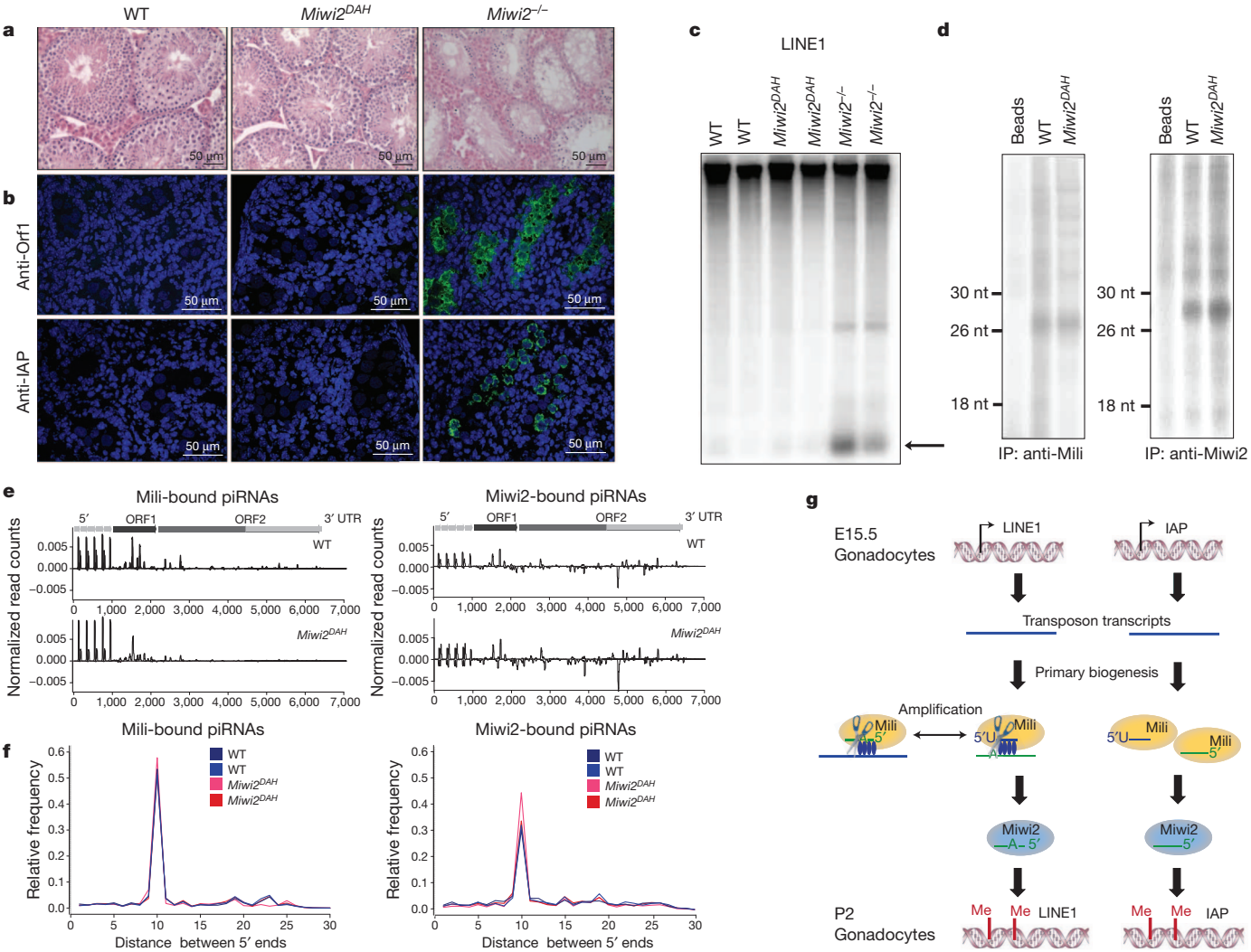


Figure 4 | Normal spermatogenesis and transposon silencing in *Miwi2*^{DAH} mice. **a**, Haematoxylin- and eosin-stained wild-type, *Miwi2*^{DAH} and *Miwi2*$^{-/-}$ testis sections from 3-month-old mice. **b**, Immunofluorescence using anti-L1 Orf1 and anti-IAP Gag antibodies (green) and DAPI-stained DNA (blue) on wild-type, *Miwi2*^{DAH} and *Miwi2*$^{-/-}$ E16.5 fetal testis sections are shown. **c**, Methylation-sensitive Southern blot on HpaII-digested DNA extracted from P10 wild-type, *Miwi2*^{DAH} and *Miwi2*$^{-/-}$ testis using a L1 promoter probe is shown. The arrow indicates the identity of the methylation-sensitive fragment. **d**, Mili (left) and Miwi2 (right) RNPs were immunoprecipitated from E16.5 fetal testis of the indicated genotypes shown as in Fig. 2a. **e**, Mapping of the Mili-bound (left) and Miwi2-bound (right) piRNAs to the consensus of L1. Positive and negative values indicate sense and antisense piRNAs, respectively. Schematic representation of L1 is shown (above). **f**, Ping-pong analysis of Mili- and Miwi2-bound piRNAs from biological replicates of wild-type and *Miwi2*^{DAH} E16.5 fetal gonadocytes are shown. The frequency of the distance between 5′ ends of complementary piRNAs for L1 is presented as in Fig. 2e. **g**, Model. L1 element silencing is dependent upon Mili's endonuclease activity for piRNA amplification. IAP silencing is dependent upon Mili and Miwi2 but independent of piRNA amplification. The small blue and green lines represent sense and antisense piRNAs, respectively.

©2011 Macmillan Publishers Limited. All rights reserved

(Fig. 4d). No qualitative or quantitative differences in piRNA origin or amplification were observed in Mili- or Miwi2-bound piRNAs from *Miwi2*^{DAH} fetal gonadocytes (Fig. 4e, f and Supplementary Figs 13–15). The lack of physiological and molecular phenotype in *Miwi2*^{DAH} mice argues against the possibility that Miwi2-mediated cleavage events functionally contribute to either slicing of nuclear transposon transcripts or piRNA amplification necessary for the establishment of transposon silencing. In support of the latter conclusion, we find piRNA amplification as judged by the existence of a robust ping-pong signature in Mili-bound piRNAs from *Miwi2*$^{-/-}$ fetal gonadocytes (Supplementary Fig. 16).

We propose a revised model of mammalian embryonic piRNA function, whereby Mili's endonuclease activity initiates secondary piRNA processing with an intra-Mili ping-pong cycle fuelling piRNA amplification (Fig. 4g). Mili's endonuclease activity is specifically required for the expansion of L1 and IAP piRNAs within Mili RNPs and the normal accumulation of all classes of piRNAs within Miwi2 RNPs. The defective piRNA pathway in *Mili*^{DAH} mice results in the failure to repress specifically L1, revealing distinct silencing requirements for the respective transposons. The establishment of L1- but not IAP-silencing is strictly dependent upon Mili's endonuclease activity and secondary piRNA biogenesis (Fig. 4g). The sufficiency of primary piRNA processing in *Mili*^{DAH} mice to direct Miwi2-mediated IAP repression illustrates fundamental differences in piRNA dosage required for silencing of the respective transposons. This difference may reflect their genomic burden—L1 occupies approximately 19% of the mouse genome whereas IAPs account for only 0.2% (ref. 28). Therefore, amplification of the L1 piRNA pool may be essential to program sufficient Miwi2 to target the prolific L1 element. In conclusion, we show that a single Piwi endonuclease supports piRNA amplification and distinct transposon silencing. It remains to be seen if these are conserved features of piRNA-mediated transposon silencing.

METHODS SUMMARY

Histology. Testes were fixed in Bouin's fixative overnight at 4 °C temperature, paraffin embedded and sectioned at 6–8-μm thickness. Sections were then stained (haematoxylin and eosin or periodic acid Schiff) by routine methods.

Immunofluorescence. For immunofluorescence, E16.5 fetal testes were freshly embedded in OCT, 6 μm sections cut and fixed in 4% paraformaldehyde. For Mili co-localization with Mvh, Tdrd1 and Mov10L1, sections were boiled twice for 5 min in 10 mM pH 6 sodium citrate solution for antigen retrieval. Sections were blocked for 30 min at room temperature in 10% normal donkey serum, 2% BSA and 0.1 M glycine. Primary antibody incubation was done overnight at 4 °C in the blocking buffer. Anti-rabbit Alexa-488-conjugated (1:1,000; Invitrogen) and anti-mouse Alexa-546-conjugated (1:1,000; Invitrogen) antibodies were used as secondary antibodies. DAPI (3 μg μl^{-1}) (Sigma) was used to stain DNA. For Miwi2, L1 Orf1 and IAP staining, paraformaldehyde-fixed sections were permeabilized with 0.1% Triton X-100 for 10 min and then blocked in TBS-T 10% normal donkey serum and processed as above. A Leica TCS SP5 confocal microscope was used to acquire images.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 31 January; accepted 8 September 2011.
Published online 23 October; corrected online 30 October 2011 (see full-text HTML version for details).

1. Ghildiyal, M. & Zamore, P. D. Small silencing RNAs: an expanding universe. *Nature Rev. Genet.* **10,** 94–108 (2009).
2. Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G. J. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316,** 744–747 (2007).
3. Carmell, M. A. *et al.* MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev. Cell* **12,** 503–514 (2007).
4. Kuramochi-Miyagawa, S. *et al.* DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev.* **22,** 908–917 (2008).
5. Brennecke, J. *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128,** 1089–1103 (2007).
6. Gunawardane, L. S. *et al.* A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in *Drosophila*. *Science* **315,** 1587–1590 (2007).
7. Aravin, A. A. *et al.* A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol. Cell* **31,** 785–799 (2008).
8. Song, J. J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305,** 1434–1437 (2004).
9. Patel, D. J. *et al.* Structural biology of RNA silencing and its functional implications. *Cold Spring Harb. Symp. Quant. Biol.* **71,** 81–93 (2006).
10. Saito, K. *et al.* Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev.* **20,** 2214–2222 (2006).
11. Bestor, T. H. & Bourc'his, D. Transposon silencing and imprint establishment in mammalian germ cells. *Cold Spring Harb. Symp. Quant. Biol.* **69,** 381–388 (2004).
12. Cheloufi, S., Dos Santos, C. O., Chong, M. M. & Hannon, G. J. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* **465,** 584–589 (2010).
13. Cifuentes, D. *et al.* A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. *Science* **328,** 1694–1698 (2010).
14. O'Carroll, D. *et al.* A Slicer-independent role for Argonaute 2 in hematopoiesis and the microRNA pathway. *Genes Dev.* **21,** 1999–2004 (2007).
15. Liu, J. *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305,** 1437–1441 (2004).
16. Maniataki, E. & Mourelatos, Z. A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes Dev.* **19,** 2979–2990 (2005).
17. Aravin, A. A. *et al.* Cytoplasmic compartmentalization of the fetal piRNA pathway in mice. *PLoS Genet.* **5,** e1000764 (2009).
18. Kojima, K. *et al.* Associations between PIWI proteins and TDRD1/MTR-1 are critical for integrated subcellular localization in murine male germ cells. *Genes Cells* **14,** 1155–1165 (2009).
19. Reuter, M. *et al.* Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. *Nature Struct. Mol. Biol.* **16,** 639–646 (2009).
20. Vagin, V. V. *et al.* Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. *Genes Dev.* **23,** 1749–1762 (2009).
21. Wang, J., Saxe, J. P., Tanaka, T., Chuma, S. & Lin, H. Mili interacts with tudor domain-containing protein 1 in regulating spermatogenesis. *Curr. Biol.* **19,** 640–644 (2009).
22. Zheng, K. *et al.* Mouse MOV10L1 associates with Piwi proteins and is an essential component of the Piwi-interacting RNA (piRNA) pathway. *Proc. Natl Acad. Sci. USA* **107,** 11841–11846 (2010).
23. Chuma, S. *et al.* Tdrd1/Mtr-1, a tudor-related gene, is essential for male germ-cell differentiation and nuage/germinal granule formation in mice. *Proc. Natl Acad. Sci. USA* **103,** 15894–15899 (2006).
24. Kuramochi-Miyagawa, S. *et al.* MVH in piRNA processing and gene silencing of retrotransposons. *Genes Dev.* **24,** 887–892 (2010).
25. Kuramochi-Miyagawa, S. *et al.* Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* **131,** 839–849 (2004).
26. Elbashir, S. M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15,** 188–200 (2001).
27. Martinez, J. & Tuschl, T. RISC is a 5′ phosphomonoester-producing RNA endonuclease. *Genes Dev.* **18,** 975–980 (2004).
28. Mandal, P. K. & Kazazian, H. H. Jr. SnapShot: Vertebrate transposons. *Cell* **135,** 192–192 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to R. Pillai, B. Cullen, S. Martin, S. Chuma and J. Lykke-Andersen for antibodies used in this study. This study was technically supported by of EMBL's genomic core facility. We are very grateful to V. Benes, R. Pillai and S. van Dongen for advice. We are grateful to M. Reuter for assistance with the preparation of immunoprecipitations for mass spectroscopy. This study was technically supported by EMBL Monterotondo's FACS and Microscopy core facilities. We are grateful to A. Wutz for A9 ES cells. We also acknowledge the services of J. Rientjes from Monash University's Gene Recombineering Facility. We are also very grateful to C. Kutter and D. Odom for advice on small RNA library generation.

Author Contributions S.D.F. contributed to the design, execution and analysis of the majority of experiments on *Mili*^{DAH} and *Miwi2*^{DAH} mice. N.B. performed the bioinformatic analysis presented in the manuscript with initial assistance from C.A.-G. M.D.G. analysed the spermatogenic defects as well as undertook the co-localization studies in the respective mouse strains. A.S. performed the bisulphite sequencing experiments. C.F. and C.A. performed the electron microscopy experiments. P.N.M. established the 8-cell embryo ES cell injection procedure. A.J.E. supervised the bioinformatic analysis. D.O'C. conceived and supervised this study and wrote the final version of the manuscript.

Author Information All raw sequencing data are deposited in ArrayExpress (accession number E-MTAB-730) and European Nucleotide Archive (ERP000778). The *Mili*^{DAH}, *Miwi2*^{DAH} and *Miwi2* null (*Miwi2*$^{-}$) alleles have been deposited at EMMA (http://www.emmanet.org/) and will be freely available on a non-collaborative basis. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.O'C. (ocarroll@embl.it).

©2011 Macmillan Publishers Limited. All rights reserved

METHODS

Mouse strains. *Mili*^{DAH} allele: the Piwi domain of Mili is encoded in exons 20 to 23. Mili aspartic acid 813, the second aspartic acid of the catalytic triad, is encoded in exon 21. To generate the *Mili*^{D813A} allele we replaced wild-type exon 21 with a mutant exon where the aspartic acid 813 codon is mutated to encode an alanine. A targeting construct was recombineered that contains homology arms and a *frt* flanked neomycin (*neo*) cassette 3′ of exon 21 that contains the *Mili*^{D813A} mutation. Southern blotting of the individual ES-cell-derived clone genomic SacI-digested DNA with a 3′ external probe was used to identify homologous recombinants. A 9.2-kb DNA fragment corresponds to the wild-type *Mili* locus; integration of the *neo*^{r}*-frt* flanked cassette 3′ of exon 21 introduces an additional *SacI* site, thus decreasing the size of the *SacI* DNA fragment recognized to 8.0 kb in the targeted allele. Flp-mediated recombination and excision of the *neo*^{r}*-frt* flanked cassette results in a 6-kb SacI DNA fragment recognized by the external 3′ probe, which is diagnostic of the *Mili*^{DAH} allele.

Miwi2^{DAH} allele: the Piwi domain of Miwi2 is encoded in exons 14–19. Miwi2 aspartic acid 761, the second aspartic acid of the catalytic triad, is encoded in exon 17. To generate the *Miwi2*^{D761A} allele we replaced wild-type exon 17 with a mutant exon where the aspartic acid 761 codon is mutated to encode an alanine. A targeting construct was recombineered that contains homology arms and a *frt* flanked neomycin (*neo*) cassette 3′ of exon 17 that contains the *Miwi2*^{D761A} mutation. Southern blotting of the individual ES-cell-derived clone genomic BsrGI-digested DNA with an external 5′ probe was used to identify homologous recombinants. An 11.8-kb DNA fragment corresponds to the wild-type *Miwi2* locus; integration of the *neo*^{r}*-frt* flanked cassette 3′ of exon 17 introduces an additional *BsrGI* site, thus decreasing the size of the *BsrGI* DNA fragment recognized to 8.4 kb. Flp-mediated recombination and excision of the *neo*^{r}*-frt* flanked cassette results in a 6.4-kb BsrGI DNA fragment recognized by the external 3′ probe, which is diagnostic of the *Miwi2*^{DAH} allele.

Miwi2 null allele: for the *Miwi2*$^{-}$ allele, we flanked exon 17 with *loxP* sites. Cre-mediated deletion of exon 17 results in out-of-frame splicing between exon 16 and 18, resulting in stop codons before the last exon, targeting the mutant transcript for nonsense-mediated decay[29]. Should a fraction of the mutant transcript escape nonsense-mediated decay, a truncated protein would be made that lacks most of the Piwi domain and thus would probably be non-functional. To generate this allele, a targeting construct was generated that contains the same homology arms as the *Mili*^{D813A} construct, an *frt* flanked *neo* cassette with a *loxP* site 3′ of exon 17 and a second 5′ *loxP* site. Southern blotting of the individual ES-cell-derived clone genomic KpnI-digested DNA with an external 5′ probe was used to identify homologous recombinants. A 10.1-kb DNA fragment corresponds to the wild-type *Miwi2* locus; integration of the *neo*^{r}*-frt2-loxP* cassette 3′ of exon 17 introduces an additional KpnI site, thus decreasing the size of the KpnI DNA fragment recognized to 8.0 kb. Cre-mediated recombination and excision of exon 17 and *neo*^{r} *loxP* flanked cassette results in a 5.3-kb KpnI DNA fragment recognized by the external 3′ probe, which is diagnostic of the *Miwi2* null (*Miwi2*$^{-}$) allele.

The Miwi2 targeting constructs and the Mili targeting construct were electroporated into IB10[30] and A9 ES cells, respectively. A9 ES cells are derived from hybrid embryos resulting from a 129/Sv male by C57BL/6 female cross (A. Wutz, manuscript in preparation). Southern blotting as described above of the individual ES-cell-derived clones was used to identify homologous recombinants. IB10-targeted ES cells were used to generate chimaeras for the respective *Miwi2* targeted alleles by standard blastocyst injections. A9-targeted ES cells were injected into C57BL/6 8-cell-stage embryos for the generation of fully ES-cell-derived mice following a procedure similar to one described previously[31] but using a Piezo Impact Unit (PMM150FU, Prime Tech) rather than a laser to puncture the zona pellucida of the host embryo. The *Mili*^{DAH} and *Miwi2*^{DAH}-targeted mice were then crossed to the FLP-expressing transgenic mice (FLPeR)[32] to remove the *frt* flanked *neo*^{r} cassette, resulting in the generation of *Mili*^{DAH} and *Miwi2*^{DAH} alleles, respectively. Mice heterozygous for the *loxP* flanked *Miwi2*-targeted allele were crossed to Deleter Cre[33] to generate the *Miwi2* null (*Miwi2*$^{-}$) allele. The mice analysed in this study were on a mixed C57BL/6 and 129 genetic background.

All of the mice were bred and maintained in EMBL Mouse Biology Unit, Monterotondo in accordance with current Italian legislation (Art. 9, 27 January 1992, number 116) under license from the Italian health ministry.

Antibodies. Rabbit polyclonal antibodies against mouse Miwi2 were generated using the same epitope as described previously[4] and used for immunoprecipitation and immunofluorescence (1:200) experiments. A mouse monoclonal antibody against Mili was obtained from R. Pillai and used for immunoprecipitation and immunofluorescence (1:1,000) experiments. The following antibodies were used at the indicated dilutions for immunofluorescence: anti-Orf1 L1 (S. Martin; 1:250), anti-IAP Gag (B. Cullen; 1:500), anti-Mvh (Abcam (ab13840); 1:200), anti-Mov10L1 (R Pillai; 1:200), anti-Tdrd1 (R. Pillai; 1:200), anti-Tdrd9 (S. Chuma; 1:200) and anti-Dcp1a (J. Lykke-Andersen; 1:500).

Histology. Testes were fixed in Bouin's fixative overnight at 4 °C temperature, paraffin embedded and sectioned at 6–8-μm thickness. Sections were then stained (haematoxylin and eosin or periodic acid Schiff) by routine methods.

Immunofluorescence. For immunofluorescence, E16.5 fetal testes were freshly embedded in OCT, 6 μm sections cut and fixed in 4% paraformaldehyde. For Mili co-localization with Mvh, Tdrd1 and Mov10l, sections were boiled twice for 5 min in 10 mM pH 6 sodium citrate solution for antigen retrieval. Sections were blocked for 30 min at room temperature in 10% normal donkey serum, 2% BSA and 0.1 M glycine. Primary antibody incubation was done overnight at 4 °C in the blocking buffer. Anti-rabbit Alexa-488-conjugated (1:,1000; Invitrogen) and anti-mouse Alexa-546-conjugated (1:1,000; Invitrogen) antibodies were used as secondary antibodies. DAPI (3 μg μl^{-1}) (Sigma) was used to stain DNA. For Miwi2, L1 Orf1 and IAP staining, paraformaldehyde-fixed sections were permeabilized with 0.1% Triton X-100 for 10 min and then blocked in TBS-T 10% normal donkey serum and processed as above.

A Leica TCS SP5 confocal microscope was used to acquire images.

Detection of apoptotic cells. Detection of apoptotic cells was performed on paraformaldehyde-fixed paraffin-embedded testis section using the *in situ* cell death detection kit (Roche) and developed with DAB substrate (Roche). Sections were also stained with haematoxylin.

Electron microscopy. Fetal testes were fixed in 2.5% glutaraldehyde in 50 mM cacodylate buffer supplemented with 2% sucrose, 50 mM KCl, 2.6 mM $CaCl_2$ and 2.6 mM $MgCl_2$ for 30 min at 4 °C and rinsed in 50 mM cacodylate buffer. Samples were incubated in 2% osmium in 50 mM cacodylate buffer for 40 min on ice, rinsed in water and incubated in 0.5% uranylacetate in water for 30 min on ice. The contrast enhancement procedure was followed by a stepwise dehydration in ethanol, up to 100% ethanol and infiltration in EPON (Roth) for embedding. Polymerization was done at 60 °C. Ultrathin sections (60–70-nm thickness) of the testis were obtained using an ultramicrotome (Leica Microsystems) and the sections were mounted on formvar-coated slot grids, and contrasted with uranylacetate and lead citrate. The sections were then viewed in a CM120 biotwin electron microscope (FEI) operating at 100 kV. Digital acquisitions were made with a Keen View CCD camera (Soft Imaging System).

Germ cell isolation. The previously described *Oct4-GFP* (ref. 34) allele was crossed into *Mili*^{DAH} and *Miwi2*^{DAH} mice to label germ cells with eGFP. To isolate E16.5 and postnatal day 7 germ cells, single cell suspensions of testis were obtained by two-step enzymatic digestion and GFP-positive FACS sorted.

CpG methylation analysis. Methylation-sensitive Southern blotting and bisulphite methylation analysis were performed as described[4].

Mass spectrometry. Mili was immunoprecipitated as described[19], resolved on gel and the entire lane with the exception of the immunoglobulin fragments was subjected to liquid chromatography coupled to tandem mass spectrometry (LC-MSMS) on a LTQ Orbitrap Velos (Thermo Fisher Scientific) instrument.

RT-qPCR analysis. Total RNA was isolated from 10 day post-partum testis using Trizol according to the manufacturer's instruction and treated with Turbo DNase RNase-free (Ambion). cDNA synthesis was performed with SuperScript III Reverse Transcriptase (Invitrogen) with Random Hexamers (Invitrogen). Quantitative PCR was carried out by using SYBR Green I Master mix (Roche) on LightCycler 480 system (Roche). Three animals for each genotype were examined and assays were always done in triplicate. Primers for qPCR were used as described[3].

Small RNA library generation. Mili and Miwi2 RNPs were immunoprecipitated as described[19]. Small RNA libraries were generated as described previously[35] but using adaptors suitable for sequencing on the Ilumina platform.

Data analysis. Sequencing data were processed from the FASTQ format and analysed using R/Bioconductor[36]. Barcode sequences were resolved by sample with no mismatches allowed, 5′/3′ adaptors were stripped, sequences were filtered for low-complexity regions and finally size-selected for reads between 24–30 nt. Processed reads for each sample were mapped against the mouse genome (NCBIm37) using Bowtie 0.12.5 (ref. 37) allowing for two mismatches and requesting all matching sites. For reads mapping to multiple distinct loci, only the first 100 were reported according to match score. Mapped reads were categorized according to genomic annotations from Ensembl[38] Mouse v58 (LTR, LINE, SINE, genic, non-coding RNA). Reads not mapping to any recorded genomic element were classed as 'other'. Counts for reads mapping to multiple loci were divided by the total number of loci. In the specific cases of LINE and IAP elements, all reads were also mapped against representative canonical sequences obtained from GenBank[39] (M13002.1, EU183301.1) using Bowtie allowing for up to three mismatches. The number of repeat mapping reads was divided by repeat length and the number of genome mapping reads per billion processed reads. Genome-wide visualizations of read mapping to the mouse genome were obtained using Circos[40] plots and were

©2011 Macmillan Publishers Limited. All rights reserved

scaled between each set of four samples according to library size (see figure legend). Line-track *y*-axis maxima are set to one-tenth of the highest peak from the library with most reads. Heat-map minima were set to one-half the average read count across all bins and maxima to one-half the maximum read count across all bins. Heat-map scaling across colour-space was performed using the 'scale_log_base = 5' parameter of Circos to sample colour space better.

For ping-pong analysis, only reads mapping to repeat elements were considered. For each pair of sense/antisense overlapping reads, the distance between their 5′ ends was recorded and counts were represented as relative frequencies within samples for each repeat element.

29. Wen, J. & Brogna, S. Nonsense-mediated mRNA decay. *Biochem. Soc. Trans.* **36,** 514–516 (2008).
30. Robanus-Maandag, E. *et al.* p107 is a suppressor of retinoblastoma development in pRb-deficient mice. *Genes Dev.* **12,** 1599–1609 (1998).
31. Poueymirou, W. T. *et al.* F0 generation mice fully derived from gene-targeted embryonic stem cells allowing immediate phenotypic analyses. *Nature Biotechnol.* **25,** 91–99 (2007).
32. Farley, F. W., Soriano, P., Steffen, L. S. & Dymecki, S. M. Widespread recombinase expression using FLPeR (flipper) mice. *Genesis* **28,** 106–110 (2000).
33. Schwenk, F., Baron, U. & Rajewsky, K. A *cre*-transgenic mouse strain for the ubiquitous deletion of *loxP*-flanked gene segments including deletion in germ cells. *Nucleic Acids Res.* **23,** 5080–5081 (1995).
34. Yoshimizu, T. *et al.* Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice. *Dev. Growth Differ.* **41,** 675–684 (1999).
35. Hafner, M. *et al.* Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods* **44,** 3–12 (2008).
36. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5,** R80 (2004).
37. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
38. Flicek, P. *et al.* Ensembl 2011. *Nucleic Acids Res.* **39,** D800–D806 (2011).
39. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. GenBank. *Nucleic Acids Res.* **36,** D25–D30 (2008).
40. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19,** 1639–1645 (2009).

©2011 **Macmillan Publishers Limited. All rights reserved**

LETTER

doi:10.1038/nature10672

Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing

Michael Reuter[1,2], Philipp Berninger[1,2], Shinichiro Chuma[3,4], Hardik Shah[5], Mihoko Hosokawa[3,4], Charlotta Funaya[6], Claude Antony[6], Ravi Sachidanandam[5] & Ramesh S. Pillai[1,2]

Repetitive-element-derived Piwi-interacting RNAs (piRNAs)[1,2] act together with Piwi proteins Mili (also known as Piwil2) and Miwi2 (also known as Piwil4) in a genome defence mechanism that initiates transposon silencing via DNA methylation in the mouse male embryonic germ line. This silencing depends on the participation of the Piwi proteins in a slicer-dependent piRNA amplification pathway and is essential for male fertility[3,4]. A third Piwi family member, Miwi (also known as Piwil1), is expressed in specific postnatal germ cells and associates with a unique set of piRNAs of unknown function[5–7]. Here we show that Miwi is a small RNA-guided RNase (slicer) that requires extensive complementarity for target cleavage *in vitro*. Disruption of its catalytic activity in mice by a single point mutation causes male infertility, and mutant germ cells show increased accumulation of LINE1 retrotransposon transcripts. We provide evidence for Miwi slicer activity directly cleaving transposon messenger RNAs, offering an explanation for the continued maintenance of repeat-derived piRNAs long after transposon silencing is established in germline stem cells. Furthermore, our study supports a slicer-dependent silencing mechanism that functions without piRNA amplification. Thus, Piwi proteins seem to act in a two-pronged mammalian transposon silencing strategy: one promotes transcriptional repression in the embryo, the other reinforces silencing at the post-transcriptional level after birth.

To investigate the biochemical role of Miwi, we affinity-purified Miwi ribonucleoprotein (RNP) complexes from mouse testes. They contained Miwi, its interaction partners[8,9] (Supplementary Table 1), and an abundant species of ~30-nucleotide (nt) small RNAs[6] (Fig. 1a and Supplementary Fig. 1a–c). Deep sequencing of Miwi-associated piRNAs from developmentally staged mouse testes revealed that Miwi is primed with a largely similar population of piRNAs in meiotic and post-meiotic germ cells (Supplementary Fig. 1d–g and Supplementary Table 2).

Some members of the Argonaute family, including Piwi proteins[10,11], are small RNA-guided nucleases (slicers) that catalyse cleavage of target nucleic acids[12]. The slicer activity depends on the presence of a catalytic motif (Asp-Asp-His; DDH motif)[13,14] in the Piwi domain of Argonautes, which is also conserved in Miwi. To biochemically confirm this activity, we incubated[15] Miwi complexes with an RNA target bearing perfect complementarity to one Miwi-associated piRNA (piR-A). The target was cleaved (Fig. 1b) at a position 10-nt downstream of the 5′ end of the guide piRNA, similar to that produced by the established slicer, human AGO2 (ref. 13; Supplementary Fig. 2d). The addition of EDTA abolished the endonucleolytic cleavage, proving the Mg^{2+} dependency of the reaction[14,16] (Fig. 1b). DNA was not recognized as a substrate (Supplementary Fig. 2c). These results confirm that Miwi is a bona fide slicer.

In agreement with crystal structures[14] of Argonaute catalytic complexes, mismatch between the first position of the piRNA and the target did not affect cleavage activity (Fig. 1c). Although mismatches at several 3′-terminal nucleotides were tolerated, base pairing of nucleotides 2–22 was required for efficient target cleavage by Miwi (Fig. 1c, d). Similar assays with a second target (for guide piR-E) gave consistent outcomes, but highlighted the need for increased complementarity to the target (Supplementary Fig. 2f). Hence, we conclude that Miwi-associated piRNAs guide endonucleolytic cleavage of target RNAs with extensive complementarity.

To determine the role of Miwi catalytic activity *in vivo*, we generated mice ($Miwi^{ADH}$) with a point mutation in Miwi at the first aspartate (D633A) of the catalytic motif (Supplementary Fig. 3), a mutation known to disrupt the activity in other slicers[13,14,17,18]. The $Miwi^{ADH}$ mutation was dominant negative, as $Miwi^{+/ADH}$ male mice were sterile. Female fertility was unaffected by this mutation. As expected, $Miwi^{+/ADH}$ testes extracts had slightly reduced slicer activity compared to the wild type (Fig. 2a). Spermatogenesis in $Miwi^{+/ADH}$ mice closely resembled that of the control mice ($Miwi^{+/-}$) until spermiogenesis[19]

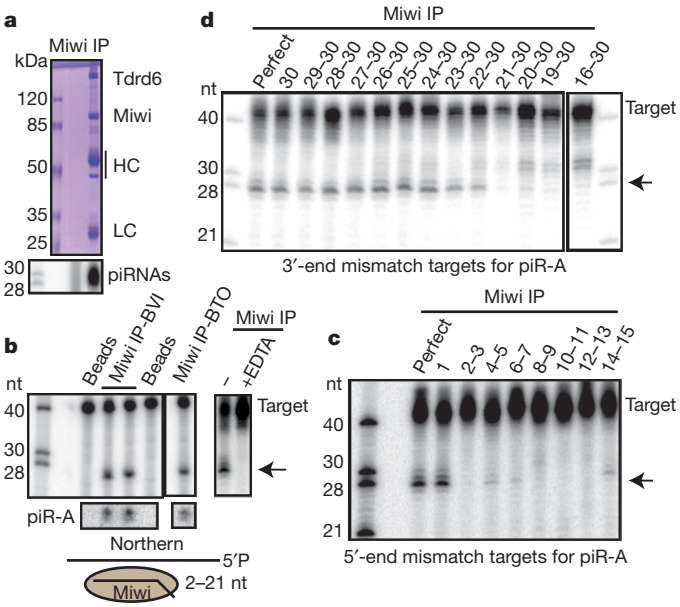


Figure 1 | Miwi is a slicer requiring extensive complementarity for target cleavage. **a**, Coomassie-stained immunopurified (BVI antibody) Miwi complexes and 5′-end-labelled associated small RNAs (piRNAs). IP, immunoprecipitates. HC and LC are antibody heavy and light chains, respectively. **b**, Slicer assays with Miwi complexes (purified with BVI or BTO antibodies) or beads alone (contol) and a 5′-end-labelled RNA target bearing complementarity to endogenous piR-A (detected by northern blotting, bottom). Addition of EDTA abolishes Miwi slicer activity. Arrow indicates the 5′ cleavage product. **c**, **d**, Slicer assays with target RNAs carrying mismatches at indicated positions relative to the guide piRNA. Protein molecular weight markers in kDa and RNA markers in nucleotides (nt) are shown.

[1]European Molecular Biology Laboratory, 6 Rue Jules Horowitz, BP 181, 38042 Grenoble, France. [2]CNRS-UJF-EMBL International Unit (UMI 3265) for Virus Host Cell Interactions (UVHCI), 38042 Grenoble, France. [3]Institute for Frontier Medical Sciences, Kyoto University, Kyoto 606-8507, Japan. [4]Institute for Integrated Cell-Material Sciences (iCeMS), Kyoto University, Kyoto 606-8501, Japan. [5]Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York 10029, USA. [6]European Molecular Biology Laboratory, EMBL Meyerhof Str. 1, 69117 Heidelberg, Germany.

©2011 Macmillan Publishers Limited. All rights reserved

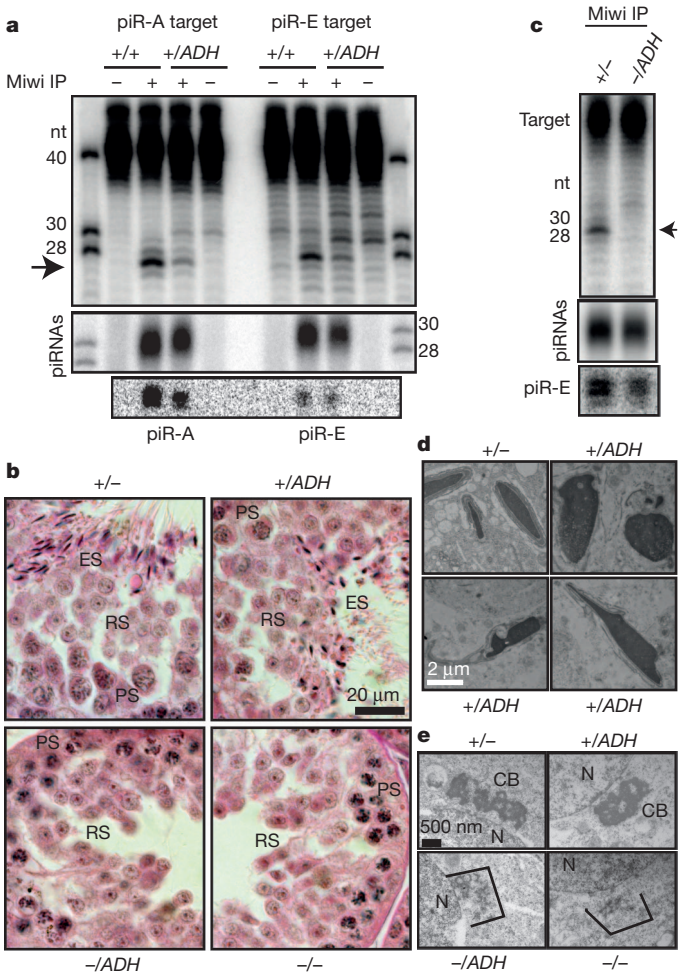


Figure 2 | Miwi catalytic activity is essential for spermatogenesis. **a**, Slicer assays with Miwi complexes from the *Miwi*$^{+/ADH}$ mutant. Total piRNAs are 5′-end labelled, whereas specific piRNAs are detected by northern blotting. **b**, Haematoxylin and eosin-stained testis sections showing the spermatogenic arrest in different indicated *Miwi* genotypes. ES, elongating spermatids; RS, round spermatids; PS, pachytene spermatocytes. **c**, Slicer assay with Miwi^{ADH} complex from the *Miwi*$^{-/ADH}$ mutant. **d**, Electron micrographs showing morphologically aberrant elongating spermatids in the *Miwi*$^{+/ADH}$ mutant. **e**, Electron micrographs of round spermatids showing the chromatoid body (CB) or its absence (residual density marked with a bracket) in *Miwi* mutants. N, nucleus.

(Fig. 2b and Supplementary Fig. 4a). Elongating spermatids are seen in *Miwi*$^{+/ADH}$ seminiferous tubules, but they exhibit abnormal nuclear morphologies (Fig. 2d) and mutant sperm were degenerated in the epididymis (Supplementary Fig. 5b). Timing and levels of chromatin-compacting protamine expression is essential for spermiogenesis[20]. Immunostaining of *Miwi*$^{+/ADH}$ mutant testis sections revealed an altered pattern for protamine 2 (Prm2), with staining detected around groups of aggregated elongating spermatid nuclei (Supplementary Fig. 6). However, the present state of our analyses does not allow precise determination of the molecular reason for the sterility observed in the *Miwi*$^{+/ADH}$ mutant.

To obtain males exclusively expressing the Miwi^{ADH} protein, we generated *Miwi*$^{-/ADH}$ mice (hereafter referred to as the catalytic mutant). Miwi slicer activity was completely abolished in the catalytic mutant (Fig. 2c and Supplementary Fig. 7b). This loss of catalytic activity was without any change in protein abundance, sub-cellular localization, arginine methylation or the piRNA-binding capacity of Miwi (Supplementary Figs 7 and 8). Similar to the *Miwi*$^{-/-}$ knockout, catalytic mutant males were sterile, with spermatogenesis being arrested after meiosis at haploid round spermatids (Fig. 2b and Supplementary Fig. 4a). Both the mutants showed defective organization of the ribonucleoprotein (RNP) granule called the chromatoid body[21] in their arrested round spermatids (Fig. 2e and Supplementary Fig. 8a). The loss of chromatoid body integrity in the catalytic mutant was without any changes in the sub-cellular localization of several Miwi-interacting proteins[8] (Supplementary Fig. 8e). Furthermore, the gene expression profiles of round spermatids were highly similar between the knockout and catalytic mutants (see later and Fig. 3d). These observations allowed us to conclude that most functions of Miwi are a result of its slicer activity.

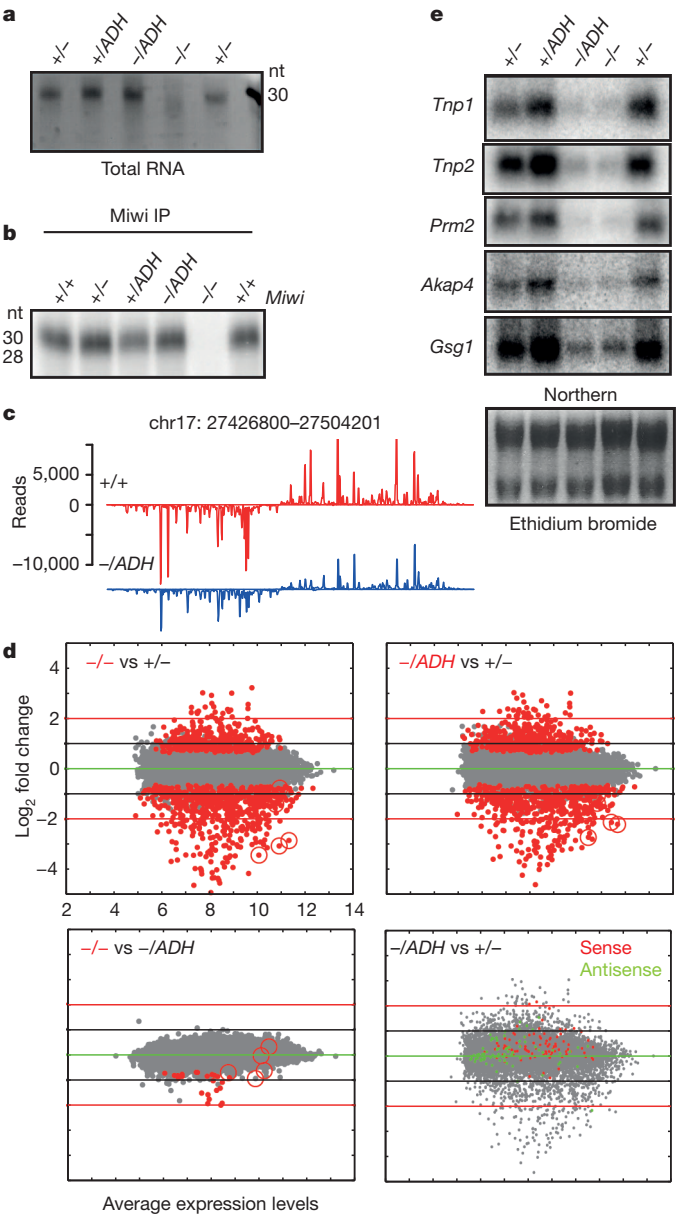


Figure 3 | Primary piRNA biogenesis is globally unaffected in the *Miwi*^{ADH} mutants and mRNAs are not targeted by piRNAs. **a**, Ethidium bromide staining of total small RNAs. **b**, Miwi-associated piRNAs detected by 5′-end labelling from indicated *Miwi* genotypes. **c**, Normalized density of Miwi piRNA reads over a genomic cluster. **d**, MA plots showing comparison of gene expression profiles in purified round spermatids from indicated *Miwi* genotypes. Significantly ($P < 0.001$) changed genes are highlighted in red. The top 100 genes having either exonic sense (red) or exonic antisense (green) piRNAs are highlighted (bottom right). **e**, Northern blot analyses of select (circled in **d**) late spermatogenesis genes and loading control (rRNA staining with ethidium bromide) are shown.

©2011 Macmillan Publishers Limited. All rights reserved

On the basis of their high preference for a uridine at the first nucleotide position (U1 bias) and their origin from the same strand within a genomic cluster, Miwi piRNAs are believed to be generated by the primary piRNA biogenesis pathway[22]. Sequencing of Miwi-associated piRNAs[6] from the *Miwi*$^{+/ADH}$ and *Miwi*$^{-/ADH}$ mutants revealed unchanged genome annotation profiles and piRNA cluster mappings, consistent with the view that primary piRNA biogenesis is independent of the slicer activity of Piwi proteins[1] (Fig. 3a–c and Supplementary Fig. 9 and Supplementary Table 2). Nevertheless, we note that the abundance of piR-A was affected in the *Miwi*$^{-/ADH}$ mutant (Supplementary Fig. 7c, d), probably owing to the developmental block[6].

The slicer activity of Miwi is reminiscent of the classical RNA interference (RNAi) action on target RNAs, which drives their rapid decay[12]. To study the effect of Miwi slicer activity on gene expression, we examined mRNA levels in purified round spermatids using microarrays (Fig. 3d and Supplementary Fig. 5c). When compared to the control (*Miwi*$^{+/-}$), several genes were affected in the knockout and catalytic mutants, some of which we confirmed by northern blot analyses (Fig. 3e). Importantly, the differences in gene expression profiles are not a result of mRNA transcripts being targeted by piRNAs (Fig. 3d and Supplementary Table 3).

To examine whether other cellular RNAs could be targeted by Miwi piRNAs, we analysed RNA-seq reads from purified round spermatids of the control (*Miwi*$^{+/-}$), and the *Miwi*$^{+/ADH}$ and *Miwi*$^{-/ADH}$ mutants (Supplementary Fig. 10). Repeat-derived piRNAs constitute ~24% of Miwi-associated sequences (Supplementary Fig. 1d), and could potentially guide Miwi slicer activity to transposon transcripts found in post-meiotic germ cells. Indeed, we observed a strong increase of LINE-specific reads in the catalytic mutant RNA-seq library (Fig. 4a and Supplementary Fig. 10c). This was quantified by quantitative polymerase chain reaction with reverse transcription (qRT–PCR) for LINE1 (L1) elements to be 7- to 22-fold higher in total RNA from *Miwi*$^{-/ADH}$ round spermatid preparations (Fig. 4b). Western blotting and immunofluorescence detection of L1ORF1p[23], a protein product expressed from active L1 elements, in the catalytic mutant round spermatids substantiated L1 upregulation (Fig. 4c, d and Supplementary Fig. 11a–c). A similar pattern of L1 upregulation was seen in the *Miwi*$^{-/-}$ knockout mutant (Fig. 4b–d). Phosphorylated histone H2AX (γ-H2AX) normally marks sites of double-stranded DNA breaks in early meiotic cells, and the XY chromosomes in pachytene spermatocytes. Coincident with L1ORF1p detection, we observed increased staining for γ-H2AX in the mutant round spermatids (Supplementary Fig. 11). This is similar to the upregulation of γ-H2AX observed in radiation-damaged round spermatids[24]. Therefore, it is clear that the resultant sterility phenotype of *Miwi* knockout and catalytic mutants is multifactorial, extending beyond the reported roles of Miwi in translation control[25] and stabilization of certain mRNAs[5].

Slicer activity generates a 5′-monophosphate-containing cleavage fragment that is complementary to the small RNA guide by 10 nt, across their 5′ ends[12,15]. We searched a global 5′-RACE (rapid amplification of 5′ complementary DNA ends) library from control (*Miwi*$^{+/-}$) round spermatids and detected an enrichment of genomic sites with 5′-RACE tags that overlap Miwi piRNAs by 10 nt (Fig. 4e and Supplementary Fig. 12). Several such overlapping RACE tags derive from LINE elements, consistent with Miwi-mediated processing of these retrotransposon mRNAs (Fig. 4e). By contrast, in the catalytic mutant, genomic sites with RACE tags overlapping across the full length of Miwi piRNAs are more abundant (peak at 29 nt; Fig. 4e), revealing the presence of uncleaved target RNAs, including those from LINE elements. The signature for slicer activity detected in the catalytic mutant is probably due to Mili[26], which is another potential slicer loaded with similar guides (Supplementary Fig. 13).

L1 silencing in embryonic[3,4] and pre-meiotic[26] male germ cells is linked to a slicer-dependent piRNA amplification pathway, a signature of which is the presence of a 10-bp overlap between the 5′ ends of

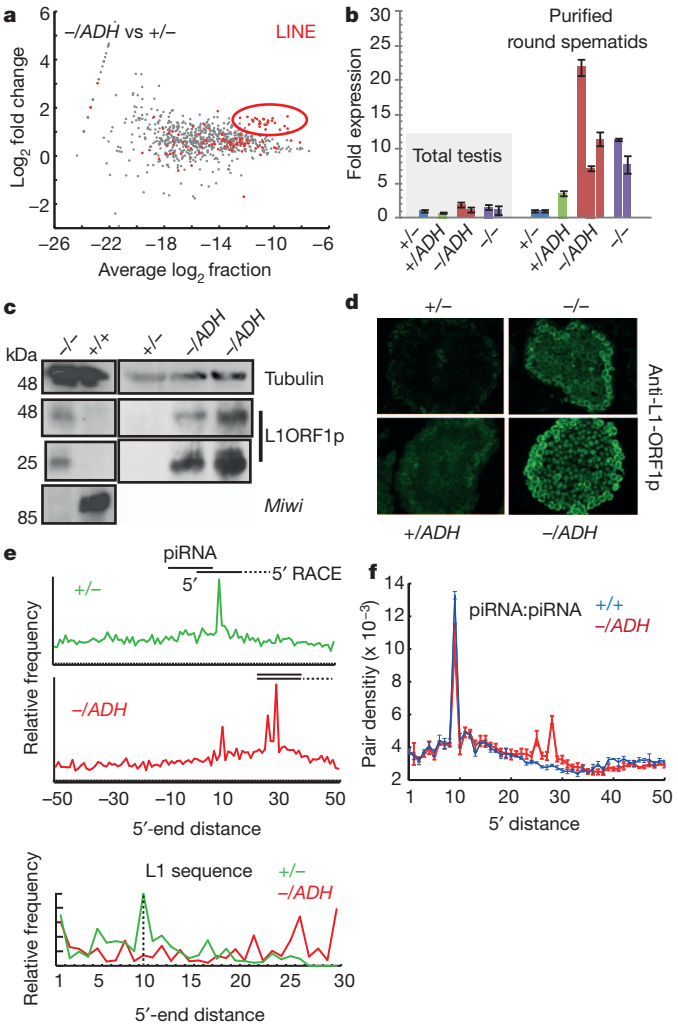


Figure 4 | Miwi piRNAs target L1 transposons. **a**, Scatter plot showing fold change of repeat reads in the RNA-seq library from the *Miwi*$^{-/ADH}$ round spermatids compared to control (+/−). LINE-specific reads are indicated in red and upregulated LINEs are circled. **b**, Quantitative RT–PCR analysis of L1 transcripts in whole testis and purified round spermatids. Each bar represents individual biological replicate and standard deviation (s.d.) between technical replicates within a biological sample is indicated. **c**, **d**, Detection of L1ORF1p in purified round spermatids by western blotting (**c**) and by immunofluorescence analysis (**d**) of testis sections. **e**, Genome-wide 5′-end overlap analysis of global 5′-RACE tags and Miwi piRNAs (top). Similar overlaps over a specific L1 sequence (bottom). **f**, 5′-end overlap analysis of Miwi piRNAs. The s.d. between three independent biological replicates is indicated.

piRNAs. We find that the *Miwi*^{ADH} mutation had no significant impact on a weak amplification signature detected in Miwi piRNA populations[27] (Fig. 4f). There was also no global change in the presence of repeat-derived piRNA reads or in DNA methylation of L1 elements in the catalytic mutant (Supplementary Table 4 and Supplementary Fig. 14). Thus, the post-transcriptional mechanism we describe here probably reinforces silencing of transposons that escape repression during spermatogenesis.

Our data indicate that a substantial portion of the biology of Miwi is implemented via its slicer activity, which in turn implies a role for its associated piRNAs. Here, we identified a role for the repeat component of Miwi piRNAs, but those deriving from unannotated genomic regions still remain a mystery. These piRNAs are not complementary to any other parts of the genome, thus the only possible targets are RNAs arising from the opposite strands of the piRNA clusters. However, we did not observe any significant upregulation of transcripts from the opposite cluster strands in the *Miwi*^{ADH} mutants. It

©2011 Macmillan Publishers Limited. All rights reserved

remains a possibility that slicer cleavage by Miwi may not always lead to RNA decay but could serve as a signal for processes involved in late spermatogenesis.

METHODS SUMMARY

Immunoprecipitations were carried out using anti-Miwi antibodies (BVI and BTO) from mouse testes extracts. Slicer assays[15] were carried out using purified Miwi complexes and 5′-end-labelled target RNAs complementary to endogenous piRNAs. The catalytic mutant mouse carries a mutation of the endogenous *Miwi* locus[5] converting the first aspartate of the DDH motif to alanine ($Miwi^{D633A}$). RNAs were deep sequenced using the Illumina platform. Total RNA from purified round spermatids was prepared using Trizol and analysed by gene microarrays and deep sequencing. A detailed description can be found in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 May; accepted 25 October 2011.
Published online 27 November 2011.

1. Siomi, M. C., Sato, K., Pezic, D. & Aravin, A. A. PIWI-interacting small RNAs: the vanguard of genome defence. *Nature Rev. Mol. Cell Biol.* **12,** 246–258 (2011).
2. Ghildiyal, M. & Zamore, P. D. Small silencing RNAs: an expanding universe. *Nature Rev. Genet.* **10,** 94–108 (2009).
3. Aravin, A. A. *et al.* A piRNA pathway primed by individual transposons is linked to *de novo* DNA methylation in mice. *Mol. Cell* **31,** 785–799 (2008).
4. Kuramochi-Miyagawa, S. *et al.* DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev.* **22,** 908–917 (2008).
5. Deng, W. & Lin, H. *miwi*, a murine homolog of *piwi*, encodes a cytoplasmic protein essential for spermatogenesis. *Dev. Cell* **2,** 819–830 (2002).
6. Girard, A., Sachidanandam, R., Hannon, G. J. & Carmell, M. A. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442,** 199–202 (2006).
7. Grivna, S. T., Beyret, E., Wang, Z. & Lin, H. A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev.* **20,** 1709–1714 (2006).
8. Vagin, V. V. *et al.* Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. *Genes Dev.* **23,** 1749–1762 (2009).
9. Vasileva, A., Tiedau, D., Firooznia, A., Muller-Reichert, T. & Jessberger, R. Tdrd6 is required for spermiogenesis, chromatoid body architecture, and regulation of miRNA expression. *Curr. Biol.* **19,** 630–639 (2009).
10. Lau, N. C. *et al.* Characterization of the piRNA complex from rat testes. *Science* **313,** 363–367 (2006).
11. Saito, K. *et al.* Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev.* **20,** 2214–2222 (2006).
12. Meister, G. & Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature* **431,** 343–349 (2004).
13. Rivas, F. V. *et al.* Purified Argonaute2 and an siRNA form recombinant human RISC. *Nature Struct. Mol. Biol.* **12,** 340–349 (2005).
14. Wang, Y. *et al.* Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* **461,** 754–761 (2009).
15. Martinez, J. & Tuschl, T. RISC is a 5′ phosphomonoester-producing RNA endonuclease. *Genes Dev.* **18,** 975–980 (2004).
16. Schwarz, D. S., Tomari, Y. & Zamore, P. D. The RNA-induced silencing complex is a Mg^{2+}-dependent endonuclease. *Curr. Biol.* **14,** 787–791 (2004).
17. Cifuentes, D. *et al.* A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. *Science* **328,** 1694–1698 (2010).
18. Cheloufi, S., Dos Santos, C. O., Chong, M. M. & Hannon, G. J. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* **465,** 584–589 (2010).
19. Caron, C., Govin, J., Rousseaux, S. & Khochbin, S. How to pack the genome for a safe trip. *Prog. Mol. Subcell. Biol.* **38,** 65–89 (2005).
20. Lee, K., Haugen, H. S., Clegg, C. H. & Braun, R. E. Premature translation of protamine 1 mRNA causes precocious nuclear condensation and arrests spermatid differentiation in mice. *Proc. Natl Acad. Sci. USA* **92,** 12451–12455 (1995).
21. Kotaja, N. *et al.* The chromatoid body of male germ cells: similarity with processing bodies and presence of Dicer and microRNA pathway components. *Proc. Natl Acad. Sci. USA* **103,** 2647–2652 (2006).
22. Lau, N. C. *et al.* Abundant primary piRNAs, endo-siRNAs, and microRNAs in a *Drosophila* ovary cell line. *Genome Res.* **19,** 1776–1785 (2009).
23. Branciforte, D. & Martin, S. L. Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. *Mol. Cell. Biol.* **14,** 2584–2592 (1994).
24. Ahmed, E. A. *et al.* Differences in DNA double strand breaks repair in male germ cell types: lessons learned from a differential expression of Mdc1 and 53BP1. *DNA Repair (Amst.)* **6,** 1243–1254 (2007).
25. Grivna, S. T., Pyhtila, B. & Lin, H. MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis. *Proc. Natl Acad. Sci. USA* **103,** 13415–13420 (2006).
26. Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G. J. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316,** 744–747 (2007).
27. Berninger, P., Jaskiewicz, L., Khorshid, M. & Zavolan, M. Conserved generation of short products at piRNA loci. *BMC Genomics* **12,** 46 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Cusack for initial support for the project and discussions. We thank D. O'Carroll and S. Khochbin for help with mouse experiments; J. Xiol and N. Nakatsuji for helpful discussions; D. O'Carroll, S. Martin, S. Kistler and R. Balhorn for antibodies. We thank the Mouse Biology Unit, European Molecular Biology Laboratory (EMBL) Monterotondo and EMBL facilities (Gene, Proteomics, Protein Expression and Purification). This work was supported by a Deutsche Forschungsgemeinschaft (DFG) postdoctoral fellowship to M.R. and grants to R.S.P. from Agence National de la Recherche (ANR) (piRmachines) and the European Union (European Research Council, ERC Starting Grant, pisilence). The R.S.P. laboratory is supported by the EMBL.

Author Contributions M.R. performed all biochemical and deep sequencing experiments. S.C., M.R. and M.H. performed imaging. C.F. and C.A. performed electron microscopy analysis. P.B., R.S., R.S.P. and H.S performed bioinformatics analysis. R.S.P., M.R., S.C. and R.S. prepared the manuscript. M.R. and R.S.P. designed research.

Author Information Deep sequencing and microarray data related to this study are deposited with the Gene Expression Omnibus under accession GSE32183. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.S.P. (pillai@embl.fr).

©2011 Macmillan Publishers Limited. All rights reserved

METHODS

Antibodies. For the production of anti-Miwi polyclonal antibodies, a GST fusion of the first 200 amino acids of Miwi was produced in *Escherichia coli* strain BL21 and injected into two rabbits (BVI and BTO) (Laboratory Animal Resource (LAR) EMBL Heidelberg). The BVI antibody was previously characterized[28]. The BTO antibody shows similar specificity to BVI in immunoprecipitations as judged by Coomassie staining of Miwi complex proteins (Supplementary Fig. 1b), size of small RNAs (Supplementary Fig. 1c), small RNA sequencing (data not shown), and by western blotting (Supplementary Fig. 7a for BVI and Fig. 4c for BTO). Both antibodies were used in immunofluorescence experiments (BVI as reported[28] and Supplementary Fig. 8b for BTO) and gave similar results.

The anti-Mili monoclonal antibody (13E-3)[28], anti-Tdrd1 (refs 28, 29), anti-Tdrd6 (ref. 30), anti-Tdrd7 (ref. 30), anti-L1 ORF1p rabbit polyclonal[23] (S. L. Martin), anti-Sm Y12 mouse monoclonal recognizing symmetrical dimethyl arginines[31] (D. Schuemperli) are described. Rabbit polyclonal antibodies to transition proteins Tnp1, Tnp2 (ref. 32) (W. S. Kistler), and protamines Prm1 and Prm2 (ref. 33) (R. Balhorn) are also described. Additional antibodies were purchased: anti-β-tubulin (Abcam), anti-γ-H2AX (Upstate/Millipore), anti-Mili (Abcam), anti-Miwi (Cell signaling technology) and anti-Mvh (Abcam).

Miwi complex purification. A single testis was dounce-homogenized in 1 ml extract buffer (10% glycerol, 50 mM Tris/HCl pH 8.0, 150 mM NaCl, 5 mM $MgCl_2$, 0.5% Na deoxycholate, 1% Triton X-100, 1× protease inhibitors (Roche), 50 μg ml^{-1} tRNA and 1× vanadyl ribonucleoside complex (Sigma)). The extract was centrifuged twice at 9,200*g*, 4 °C, 15 min. Anti-Miwi or -Mili antibodies were bound to Protein G Sepharose overnight with gentle rocking at 4 °C in IP buffer (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM $MgCl_2$). After washes in IP buffer, extract was added and incubation continued for 2–5 h at 4 °C. After washes, complexes were used in slicer assays. Bound piRNAs were extracted by proteinase K treatment, phenol-choloroform extraction and further visualized by 5′-end labelling.

For Miwi complex purification (Fig. 1a and Supplementary Fig. 1b), affinity purified anti-Miwi antibodies (25–100 μg BVI or BTO) were chemically cross-linked to protein G sepharose using dimethyl pimelimidate (DMP; Sigma D-8388). For proteomic analysis the beads-bound material was eluted with 0.1 M glycin pH 2.5–3.0, 150 mM NaCl, and after methanol precipitation, separated by SDS–PAGE.

After staining with Coomassie blue, all visible bands and intervening gel spaces were excised for identification by liquid chromatography–mass spectrometry (LC/MS/MS) (Proteomics Core Facility, EMBL Heidelberg). Briefly, after tryptic digest, peptides were separated by liquid chromatography using the nanoAcquity UPLC system (Waters) fitted with a nanoAcquity Symmetry C_{18} trapping and analytical column. Peptides were directly fed into a LTQ Orbitrap Velos (Thermo Fischer Scientific). Software MaxQuant was used for filtering the data and its preparation for MASCOT search. An unedited table with all identified proteins and the number of peptides determined is provided (Supplementary Table 1). Specificity of the identified proteins was determined by control purifications with protein G sepharose beads alone or protein G coupled to rabbit IgG (anti-HA). Some of the identified proteins, such as tudor proteins (Tdrd1 and Tdrd6)[8,9] and kinesin motor protein Kif17 (ref. 34) were already identified in previous proteomic and biochemical studies to interact with Miwi, whereas the presence of others, such as Mili, in this list might be due to their co-localization in the chromatoid body[8,21].

Slicer assays. A synthetic 40-nt RNA (for piR-A) or *in vitro* transcribed RNAs (for piR-A and piR-E), both bearing complementarity to endogenous piRNAs, were used as targets (Supplementary Fig. 2). Target RNAs were gel eluted from 15% urea-polyacrylamid gels and 5′-end labelled.

Miwi was immunopurified from testis extract (1 testis per ml buffer, ~0.5 μg anti-Miwi antibody bound to ~10 μl protein G sepharose). After the complete removal of wash buffer, beads were pre-incubated for >10 min in 10 μl slicer buffer[15]. (10 mM Tris/HCl pH 7.5, 100 mM KCl, 2 mM $MgCl_2$, 40 U RiboLock (Fermentas)) on ice. The assay was started by adding 10 μl 5′-labelled (10–50 fmol), denatured RNA (55 °C for 5–10 min followed by immediate transfer to ice). The reaction was performed in a thermoshaker (Eppendorf) for 0.5–2 h at 35 °C with constant changing modes of shaking (1,000 r.p.m., 10 s) and pause (90 s). The reaction was stopped by proteinase K treatment followed by phenol-chloroform extraction and precipitation. RNA was resolved by 15% urea–PAGE. Gels were dried and exposed to Storage Phosphor screens (GE Health).

We tested 14 different piRNAs by northern blotting to judge abundance and to identify suitable guides for slicer assays (data not shown). Of these only piR-A and piR-E guides gave detectable signals in the slicer assays reported here. So it remains a possibility that not all Miwi piRNPs are competent for slicer activity.

Northern blot. Northern analysis of small RNAs was performed on 2.5–10 μg total RNA or Miwi-associated small RNAs separated by 15% urea–PAGE. When required (Supplementary Fig. 7c), RNAs were first treated by beta-elimination[35] before PAGE separation. Small RNAs were detected with the following 5′-labelled DNA probes: piRNA-A, AAAGCTATCTGAGCACCTGTGTTCATGTCA; piRNA-E, ACCAACTCTGTCTTCAGAATATACATCAACT; and miRNA-16, GCCAATATTTACGTGCTGCTA. Probes for the detection of longer transcripts were generated from DNA fragments or PCR products by random priming (Roche).

***Miwi*^{ADH} and *Miwi* null mice.** The catalytic triad of Argonaute proteins is constituted by three amino acids (DDH motif)[13,14,36]. The first aspartate of this motif in Miwi is amino acid number 633, coded in exon 17. We mutated D633 to alanine (D633A) by gene targeting in BRUCE4 embryonic stem (ES) cells[37]. We replaced the exons 16 and 17 with a fragment containing exon 16 and a modified exon 17 with an alteration in a single codon that generates the ADH (D633A) point mutation. An *frt*-flanked neomycin (*Neo*) resistance cassette and a *loxP* site, along with an EcoRI site were placed 3′ of *Miwi*[5] exon 17. Another *loxP* site was placed 5′ of exon 16. These modifications to the *Miwi* locus are termed the *Miwi* targeted (*Miwi*^{t}) allele (Supplementary Fig. 3). Southern blotting with two independent *Miwi* probes and EcoRI-digested genomic DNA from different ES cell clones confirmed proper integration into the targeted locus (data not shown). The modified ES cells were used by standard techniques to generate *Miwi*$^{+/t}$ mice (Transgenic Mouse Facility, EMBL Monterotondo) and Southern blotting confirmed the existence of the recombined loci in targeted mice (Supplementary Fig. 3). Presence of a single 24.6-kb fragment with both probes identifies the wild-type allele. Owing to the integration of an additional EcoRI site, the 5′ probe identifies 12.6 kb whereas the 3′ probe identifies 14-kb fragments from the targeted allele.

Excision of *frt*-flanked neomycin resistance cassette to generate the *Miwi*^{ADH} knock-in allele was achieved by crossing *Miwi*$^{+/t}$ mice with the transgenic FLPeR[38] mice that ubiquitously express the Flp recombinase. To generate the *Miwi* null (*Miwi*$^{-/-}$) allele, *Miwi*$^{+/t}$ mice were crossed with Deleter Cre[39] mice that ubiquitously express the Cre recombinase, which excised the neomycin resistance cassette and exons 16 and 17 that were flanked by the inserted *loxP* sites. It most likely results in a transcript containing a premature stop codon, leading to mRNA decay via NMD[40]. We did not detect any Miwi protein or truncated versions in our *Miwi*$^{-/-}$ mice by western blotting (Supplementary Fig. 7a). All mice were maintained in the C57/BL6 genetic background.

Females carrying the *Miwi* null and ADH alleles in all combinations are fertile. We intercrossed *Miwi*$^{+/-}$ males with *Miwi*$^{+/ADH}$ females to generate *Miwi*$^{-/ADH}$ mice, which are referred to as catalytic mutant. PCR amplification of the relevant DNA region from the different genotypes identified the presence of the correct GAC (D633) to GCC (D633A) mutation in exon 17 in the *Miwi*^{ADH} mutants (Supplementary Fig. 3c). All of the mice were bred and maintained in the EMBL Mouse Biology Unit, Monterotondo in accordance with current Italian legislation under license from the Italian health ministry.

Immunofluorescence analysis. Immunofluorescence analysis[28] was performed on paraffin or OCT (Tissue-Tek Cryo-OCT compound, Fischer Scientific) embedded testes sections. Primary antibodies (1:50–1:500) and Alexa-488- or 555-coupled secondary antibodies (Invitrogen) were diluted (1:5,000) before use. DNA was stained with Hoechst 33258 and slides were mounted with Vectashield (Vector Labs). PNA rhodamine (Vector Labs) was purchased. Anti-γ-H2AX (Upstate) and mouse monoclonal γ-H2AX peptide (phospho-S139, Abcam) were purchased. The antibody (1 μg) and peptide (10 times the antibody weight) was pre-incubated overnight at 4 °C, after which the immunostaining procedure continued as usual. Samples were examined by fluorescence microscopy (Olympus BX61) and images captured using CCD (Olympus DP70). Miwi localization in the different genotypes was examined using a Leica TCS SP2 AOBS, inverted microscope. The *Tdrd9* mutant was described previously[41].

Global 5′-RACE library. Total RNA from purified round spermatids (*Miwi*$^{+/-}$ and *Miwi*$^{-/ADH}$) was depleted of rRNAs using the Ribo-Zero kit (Epicentre) and then ligated to the Illumina 5′RNA adaptor in a 10 μl reaction (10 U RNA ligase, 1 mM ATP, 0.75 μM 5′-adaptor, 20 U RNase inhibitor) at 20 °C overnight. Only RNAs carrying a natural 5′ monophosphate should be ligated because a kinase step was omitted[42]. RNAs were re-purified using the Absolutely RNA kit (Stratagene). This step allows the removal of a majority of short RNAs (<100 nt, piRNAs, adapters) from the RNA sample. The eluted RNA was phenol-chloroform extracted and precipitated. Reverse transcription was performed in a 20 μl reaction with 100 pmol random primer (5′-ACAAGCAGAAGACGGCATACGA-N_6, where N is a random nucleotide), 0.5 mM dNTP, 1× RT buffer, 5 mM DTT, 20 U RNase inhibitor and 200 U reverse transcriptase. The reaction mix was first incubated at 25 °C for 2 min. After addition of the enzyme, the reaction was incubated for 5 min at 25 °C, followed by a step at 50 °C for 40 min and 55 °C for 20 min. 50% of the RT reaction was used for PCR (98 °C, 30 s; 35 cycles: 98 °C, 10 s; 60 °C, 30 s; 72 °C, 15 s; final amplification: 72 °C, 5 min). PCR reaction was carried out with 1 U of Phusion High-Fidelity DNA Polymerase, 0.2 mM dNTP,

©2011 Macmillan Publishers Limited. All rights reserved

1.5 mM $MgCl_2$ and 250 μM primer. PCR products of ~200 bp were gel-eluted for sequencing (105 cycles) by the Illumina Genome Analyser (EMBL Gene core facility).

Small RNA and RNA-seq library preparation. Miwi-associated piRNAs were isolated from testes of wild-type mice of different ages: 14, 20 days after birth (P14, P20); and from adult mice (>P60). Adult males were used for the generation of libraries from the $Miwi^{+/ADH}$ and $Miwi^{-/ADH}$ mutant. Triplicates from independent biological samples were prepared for the catalytic mutant $Miwi^{-/ADH}$. One library was made from the $Miwi^{+/ADH}$ mutant. Libraries were prepared by sequential ligation of 5′- and 3′-RNA adapters. Products were gel eluted after each ligation step. Sequencing was for 36 cycles using the Illumina Genome Analyser IIx (EMBL Gene core facility).

Round spermatids were purified from $Miwi^{+/-}$, $Miwi^{-/ADH}$ and $Miwi^{+/ADH}$ adult animals using a BSA gradient[43]. We note that in such a purification protocol, the control (+/−) will contain the full spectrum of round spermatid stages, whereas the mutants will have only those until the respective arrested stages. Total RNA was extracted using Trizol and depleted of abundant RNAs like rRNAs using Ribo-Zero kit (Epicentre). Strand-specific RNA-seq library was prepared using the ScriptSeq mRNA-seq Library Preparation Kit (Epicentre; Cat no. SS10906) and sequenced (EMBL Gene core facility, Heidelberg). We obtained ~20 million reads for each genotype from 76-cycle Illumina runs.

Bioinformatics analyses. Reads from piRNA libraries were mapped to the genome and only perfect mappers were selected for further analysis. In-house bioinformatics tools developed by the R.S. laboratory and used in this study are described[44]. For cluster analysis, reads were mapped on to 2-kb tiles with minimum of 50 hits and moved in steps of 1 kb. The top clusters from such an analysis are shown for all the Miwi libraries (Supplementary Table 2).

5′-end overlap analysis. The three wild type (+/−) and the three catalytic mutant (−/ADH) piRNA libraries were read-depth normalized and 27–31-nt-long perfect and unique mapping reads were used in the analysis. Using a window of 300 nt, we calculated genome-wide, the correlation of 5′ ends from opposite strands with the pairs function, as previously described[27]. Afterwards, each obtained correlation profile was normalized by dividing through the total correlation measured within 300 nt and the mean and s.d. for the three +/− and three −/ADH combined correlation profiles were determined. The genomic loci giving rise to a predominant peak at 10 was investigated further. This is indicative of the presence of reads with their 5′ ends overlapping by 10 nt. The genomic annotation of those loci was determined and found to be very similar to that seen for Miwi piRNAs—with a large component of unannotated reads (data not shown).

We also investigated the 5′-end overlap distances of piRNAs with the 5′ ends of reads from the RACE library genome-wide (Fig. 4e and Supplementary Fig. 12). Here, we calculated the number of sites carrying a particular 5′ distance, as previously described[27]. We also calculated the overlap over a LINE element (L1Md_F2; chr2:69191898-69194424) that is upregulated in the catalytic mutant (Fig. 4e and Supplementary Fig. 10c).

Microarray analysis. Round spermatids were purified from $Miwi^{+/-}$, $Miwi^{-/ADH}$ (catalytic mutant), $Miwi^{-/-}$ and $Miwi^{+/ADH}$ adult mice using a BSA gradient method[43]. Gene expression profiles were analysed by Affymetrix gene array, in biological duplicates (EMBL Gene Core, Heidelberg). Microarray analysis was carried out in R/Bioconductor[45]. First, the microarrays were corrected for background with RMA and quantile-normalized with the aroma package[46]. Duplicates were combined and significant ($P < 0.001$) changing genes were identified with limma[47]. To remove unexpressed genes from the further analysis, genes having $\log_2$ intensity higher than 5.6 in at least one of the four sets were selected as expressed genes.

RNA-seq analysis. All mapped reads were normalized to an artificial library size of 1 million reads, in order to allow comparisons across libraries. For the read-depth normalized RNA-seq libraries, we determined the expression of genes by adding up all unique mapping reads falling into the exonic region of a particular gene. Genes where considered to be expressed when they carried at least 5 reads per million in one of the libraries. For these genes the fold changes were determined.

28. Reuter, M. *et al.* Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. *Nature Struct. Mol. Biol.* **16,** 639–646 (2009).
29. Chuma, S. *et al. Tdrd1/Mtr-1*, a *tudor*-related gene, is essential for male germ-cell differentiation and nuage/germinal granule formation in mice. *Proc. Natl Acad. Sci. USA* **103,** 15894–15899 (2006).
30. Tanaka, T. *et al.* Tudor domain containing 7 (*Tdrd7*) is essential for dynamic ribonucleoprotein (RNP) remodeling of chromatoid bodies during spermatogenesis. *Proc. Natl Acad. Sci. USA* **108,** 10579–10584 (2011).
31. Brahms, H. *et al.* The C-terminal RG dipeptide repeats of the spliceosomal Sm proteins D1 and D3 contain symmetrical dimethylarginines, which form a major B-cell epitope for anti-Sm autoantibodies. *J. Biol. Chem.* **275,** 17122–17129 (2000).
32. Heidaran, M. A., Showman, R. M. & Kistler, W. S. A cytochemical study of the transcriptional and translational regulation of nuclear transition protein 1 (TP1), a major chromosomal protein of mammalian spermatids. *J. Cell Biol.* **106,** 1427–1433 (1988).
33. Stanker, L. H., Wyrobek, A. & Balhorn, R. Monoclonal antibodies to human protamines. *Hybridoma* **6,** 293–303 (1987).
34. Kotaja, N., Lin, H., Parvinen, M. & Sassone-Corsi, P. Interplay of PIWI/Argonaute protein MIWI and kinesin KIF17b in chromatoid bodies of male germ cells. *J. Cell Sci.* **119,** 2819–2825 (2006).
35. Kirino, Y. & Mourelatos, Z. Mouse Piwi-interacting RNAs are 2′-O-methylated at their 3′ termini. *Nature Struct. Mol. Biol.* **14,** 347–348 (2007).
36. Qi, Y. *et al.* Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* **443,** 1008–1012 (2006).
37. Kontgen, F. & Stewart, C. L. Simple screening procedure to detect gene targeting events in embryonic stem cells. *Methods Enzymol.* **225,** 878–890 (1993).
38. Farley, F. W., Soriano, P., Steffen, L. S. & Dymecki, S. M. Widespread recombinase expression using FLPeR (flipper) mice. *Genesis* **28,** 106–110 (2000).
39. Schwenk, F., Baron, U. & Rajewsky, K. A *cre*-transgenic mouse strain for the ubiquitous deletion of *loxP*-flanked gene segments including deletion in germ cells. *Nucleic Acids Res.* **23,** 5080–5081 (1995).
40. Wen, J. & Brogna, S. Nonsense-mediated mRNA decay. *Biochem. Soc. Trans.* **36,** 514–516 (2008).
41. Shoji, M. *et al.* The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. *Dev. Cell* **17,** 775–787 (2009).
42. Karginov, F. V. *et al.* Diverse endonucleolytic cleavage sites in the mammalian transcriptome depend upon microRNAs, Drosha, and additional nucleases. *Mol. Cell* **38,** 781–788 (2011).
43. Pivot-Pajot, C. *et al.* Acetylation-dependent chromatin reorganization by BRDT, a testis-specific bromodomain-containing protein. *Mol. Cell. Biol.* **23,** 5354–5365 (2003).
44. Olson, A. J., Brennecke, J., Aravin, A. A., Hannon, G. J. & Sachidanandam, R. Analysis of large-scale sequencing of small RNAs. *Pac. Symp. Biocomput.* 126–136 (2008).
45. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5,** R80 (2004).
46. Bengtsson, H., Wirapati, P. & Speed, T. P. A single-array preprocessing method for estimating full-resolution raw copy numbers from all Affymetrix genotyping arrays including GenomeWideSNP 5 & 6. *Bioinformatics* **25,** 2149–2156 (2009).
47. Smyth, G. K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3,** article 3 (2004).

©2011 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature10577

Atomic–resolution dynamics on the surface of amyloid–β protofibrils probed by solution NMR

Nicolas L. Fawzi[1], Jinfa Ying[1], Rodolfo Ghirlando[2], Dennis A. Torchia[3] & G. Marius Clore[1]

Exchange dynamics between molecules free in solution and bound to the surface of a large supramolecular structure, a polymer, a membrane or solid support are important in many phenomena in biology and materials science. Here we present a novel and generally applicable solution NMR technique, known as dark-state exchange saturation transfer (DEST), to probe such exchange phenomena with atomic resolution. This is illustrated by the exchange reaction between amyloid-β (Aβ) monomers and polydisperse, NMR-invisible ('dark') protofibrils, a process of significant interest because the accumulation of toxic, aggregated forms of Aβ, from small oligomers to very large assemblies, has been implicated in the aetiology of Alzheimer's disease[1–6]. The ^{15}N-DEST experiment imprints with single-residue-resolution dynamic information on the protofibril-bound species in the form of ^{15}N transverse relaxation rates (^{15}N-R_2) and exchange kinetics between monomers and protofibrils onto the easily observed two-dimensional ^{1}H–^{15}N correlation spectrum of the monomer. The exchanging species on the protofibril surface comprise an ensemble of sparsely populated states where each residue is either tethered to (through other residues) or in direct contact with the surface. The first eight residues exist predominantly in a mobile tethered state, whereas the largely hydrophobic central region and part of the carboxy (C)-terminal hydrophobic region are in direct contact with the protofibril surface for a significant proportion of the time. The C-terminal residues of both Aβ40 and Aβ42 display lower affinity for the protofibril surface, indicating that they are likely to be surface exposed rather than buried as in structures of Aβ fibrils[7–10], and might therefore comprise the critical nucleus for fibril formation[11,12]. The ^{15}N$-R_2^{tethered}$ values, however, are significantly larger for the C-terminal residues of Aβ42 than Aβ40, which might explain the former's higher propensity for rapid aggregation and fibril formation[13,14].

We recently reported the existence of a dynamic exchange process under pseudo-equilibrium conditions between monomeric and very high molecular mass, soluble species of Aβ40 (ref. 15). The latter form spontaneously in concentrated (150–300 μM) unstirred aqueous solution, and comprise polydisperse, non-fibrillar aggregates typical of worm-like protofibrils (Supplementary Fig. 1a). Similar protofibrils are formed by Aβ42 under the same conditions (Supplementary Fig. 1b). These protofibrils, which we have characterized extensively by atomic force and electron microscopy (Supplementary Fig. 1), analytical ultracentrifugation (Supplementary Fig. 3) and dynamic light scattering (Supplementary Fig. 4) are far too large (1.8–85 MDa) to detect directly using solution NMR spectroscopy as their resonances are broadened beyond detection[15], and their heterogeneous nature makes biophysical characterization challenging[5,16–19]. Both the Aβ40 and Aβ42 protofibrils present in the equilibrated concentrated samples react strongly with the OC polyclonal antibody that recognizes so-called 'fibrillar' oligomers and fibrils[3] (Supplementary Fig. 6b); the A11 polyclonal antibody recognizes so-called 'pre-fibrillar' oligomers[3] and reacts with both Aβ40 and Aβ42 protofibrils, albeit weakly with the former (Supplementary Fig. 6a). The predominantly unstructured, intrinsically disordered, monomeric Aβ peptides[20,21], however, give rise to excellent solution NMR spectra[15] (Supplementary Fig. 7), and analytical ultracentrifugation indicates unambiguously that the NMR visible species is monomeric, and that there are no measurable quantities of low molecular mass oligomers present in either the dilute (Supplementary Fig. 2) or concentrated (Supplementary Fig. 5) samples. Indeed, in the concentrated samples containing monomer and protofibrils, there is no evidence by analytical ultracentrifugation of any species other than monomer and protofibrils with M_r in excess of 1.8 MDa (Supplementary Figs 4 and 5). In this work, we probe the protofibril-bound species at a residue-by-residue level using ^{15}N-DEST, a novel two-dimensional transfer-of-saturation NMR experiment that exploits the exchange process between NMR invisible ('dark state') and easily observed monomeric states.

The essence of the two-dimensional ^{15}N-DEST experiment is based on the dramatic reduction in rates of re-orientation upon binding to the surface of the protofibrils resulting in transverse relaxation rates (^{15}N-R_2) that are orders of magnitude larger than those of the free monomer. These large ^{15}N-R_2 values preclude direct observation by solution NMR, but allow for efficient partial saturation of ^{15}N longitudinal magnetization of the protofibril-bound form by a weak radio-frequency field, even at large offsets where the free monomer magnetization is completely unaffected. In the DEST experiment, ^{15}N longitudinal magnetization of the free monomer is prepared (INEPT and refocused INEPT transfer), transferred by chemical exchange to the protofibril-bound form (binding), partly saturated (weak ^{15}N saturation at a series of offsets) and transferred back to the free monomer (dissociation) where the partial saturation is recorded as an attenuation of the cross-peaks of the NMR observable free monomer relative to that in a reference spectrum obtained without saturation (see Methods and Supplementary Fig. 8).

NMR measurements were made at 10 °C, pH 6.8 using two concentrations of Aβ: a low concentration sample (50 μM) at which both Aβ40 and Aβ42 remain nearly entirely monomeric (>90%) for a long period of time (>2 months for Aβ40 and approximately 1 week for Aβ42) based on cross-peak intensities in a ^{1}H–^{15}N heteronuclear single quantum coherence correlation spectrum; and a higher concentration sample in which the fraction of monomer decreases rapidly (over approximately 2 days for Aβ42 and 1 week for Aβ40) after which a stable pseudo-equilibrium between monomer and protofibrils is maintained for more than a week[15] (Supplementary Fig. 9). The amount of protofibrils formed at a given initial concentration of Aβ is significantly higher for Aβ42 than Aβ40 (Supplementary Fig. 9). Thus, for 270 μM Aβ40 and 160 μM Aβ42 samples, the concentration of monomer decreases to pseudo-steady-state levels of approximately 110 and 35 μM, respectively.

For the monomeric Aβ40 sample, the ^{1}H–^{15}N cross-peaks in the DEST experiment display an extremely sharp saturation profile with

[1]Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, USA. [2]Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, USA. [3]National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland 20892-0520, USA.

©2011 Macmillan Publishers Limited. All rights reserved

no detectable attenuation for any residue until the saturation pulse is less than 2 kHz from the ^{15}N resonance frequencies of the monomer (Fig. 1a, circles). This behaviour, replicated for the monomeric Aβ42 sample, matches the predicted saturation profile for a ^{15}N spin with relaxation parameters measured for the monomeric species (Fig. 1a, solid line). In contrast, the ^{1}H–^{15}N cross-peaks in the 270-μM Aβ40 sample show significant partial saturation at offsets far from the monomer ^{15}N resonances. Most importantly, the saturation profiles are clearly residue specific (Fig. 1b–d, circles). Relatively narrow profiles are observed for residues in the amino(N)-terminal region (Fig. 1b), and very broad profiles for residues in the central (residues 17–21, for example, Fig. 1c) and part of the C-terminal (residues 30–40) hydrophobic regions. Intermediate profiles are observed for residues in the hydrophilic region spanning residues 25–29 (Fig. 1d) as well as for the three C-terminal residues. Qualitatively similar behaviour is observed for the 160-μM Aβ42 sample (Supplementary Fig. 10a–d). The observed residue-specific saturation transfer profiles indicate that, when bound to the surface of the protofibrils, the average ^{15}N-R_2 for each residue is highly sequence-position dependent.

The ^{15}N-DEST profiles can be quantitatively interpreted by incorporating a model of equilibrium chemical exchange within the framework of the McConnell equations[22,23]. The simplest model describing the equilibrium between free monomer ($M_{monomer}$) and peptide transiently bound to the surface of the protofibril ($M_{oligomer}$) is a pseudo-first-order process characterized by global first-order apparent association (k_{on}^{app}) and dissociation (k_{off}) rate constants (Fig. 2a) and residue-specific ^{15}N-R_1 (longitudinal) and R_2 (transverse) values for the free monomer and protofibril-bound states. The value of the k_{on}^{app} can be estimated from the difference, ΔR_2, in ^{15}N-R_2 values between the samples with protofibrils present and the monomeric reference sample (50 μM)[15]. Because transverse magnetization for the protofibril-bound states decays extremely rapidly, the maximum observed ΔR_2 value, ΔR_2^{max}, is equal to k_{on}^{app} under pseudo-equilibrium conditions (note that the observed ΔR_2 values are independent of nucleus and magnetic field[15]; see also Supplementary section 5 for a simple explanation of ΔR_2 and DEST experiments and their interpretation). The saturation profiles at two radio-frequency field strengths (170 and 350 Hz) and the ΔR_2 data for all residues (Fig. 1e for Aβ40 and Supplementary Fig. 10e for Aβ42) were fitted simultaneously, optimizing the values of k_{on}^{app} and k_{off}, and the residue-specific ^{15}N-R_2 values for the protofibril-bound species (see Methods). The ^{15}N-R_1 and R_2 values for the monomer were taken from relaxation data recorded on the low concentration (50 μM) sample. Although the ^{15}N-R_1 values for the protofibril-bound state are unknown, these ^{15}N-R_1 values have no effect on the fitted parameters within their uncertainties over the expected range ($2\,s^{-1} \geq {}^{15}$N-$R_1 \geq 0.01\,s^{-1}$) owing to the design of the experiment (see Methods). The simple two-state model fails to account simultaneously for all the experimental data in the case of both Aβ40 (Fig. 1b–e) and Aβ42 (Supplementary Fig. 10), with systematic deviations for the DEST profiles (Fig. 1b–d, dashed lines) and failure to reproduce ΔR_2 as a function of sequence (Fig. 1e, grey open circles). Extension to a three-site exchange model that adds an intermediate species either on- or off-pathway, introducing additional unknown rate constants connecting monomer, intermediate and protofibril-bound states and residue-specific ^{15}N-R_2 values for the intermediate state, also fails to fit the experimental data.

The above results led us to investigate a simple modification of the two-state model in which the ensemble of protofibril-bound states is divided into two subsets where a given residue is either in direct contact

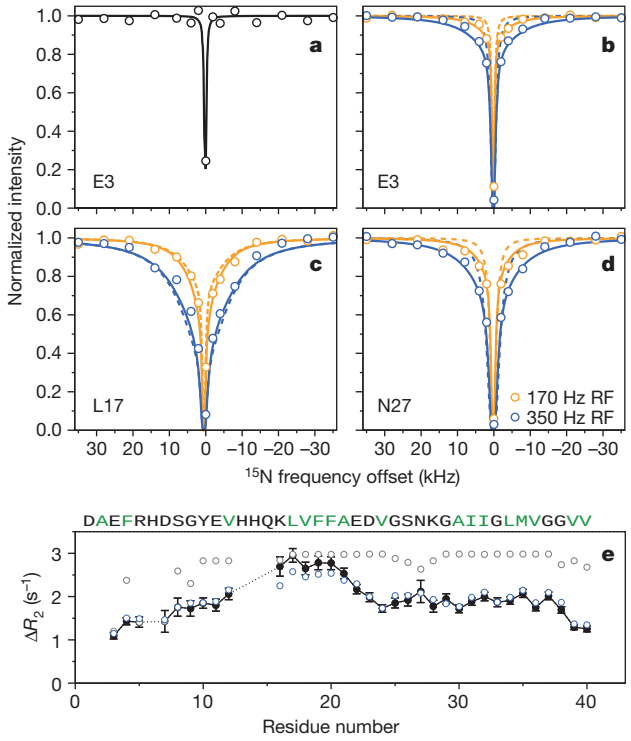


Figure 1 | ^{15}N DEST and ΔR_2 for Aβ40. Examples of normalized cross-peak intensities as a function of frequency offset from the ^{15}N carrier (118.5 p.p.m.): **a**, Glu 3 in the 50 μM Aβ40 sample (radio-frequency field strength of ^{15}N saturation pulse = 170 Hz); **b**, **c** and **d**, Glu 3, Leu 17 and Asn 27, respectively, in the 270 μM Aβ40 sample with ^{15}N saturation at radio-frequency field strengths of 170 (orange circles) and 350 (blue circles) Hz. The solid line in **a** is the calculated saturation profile for a ^{15}N spin with the experimentally determined relaxation rates for monomeric Aβ40. The dashed and solid lines in **b**–**d** are the best-fits to an exchange model with a single protofibril-bound state (cf. Fig. 2a) and to a model incorporating residues tethered and in direct contact with the protofibril (cf. Fig. 2b), respectively. The s.d. of the DEST experimental data points is approximately equal to the size of the circles. **e**, Observed (black filled circles; error bars, 1 s.d.) compared with calculated ΔR_2 values for the first (grey circles) and second (blue circles) models. The reduced χ^2 for the simultaneous best-fit to the DEST and ΔR_2 data are 9.0 for the first model and 1.8 for the second.

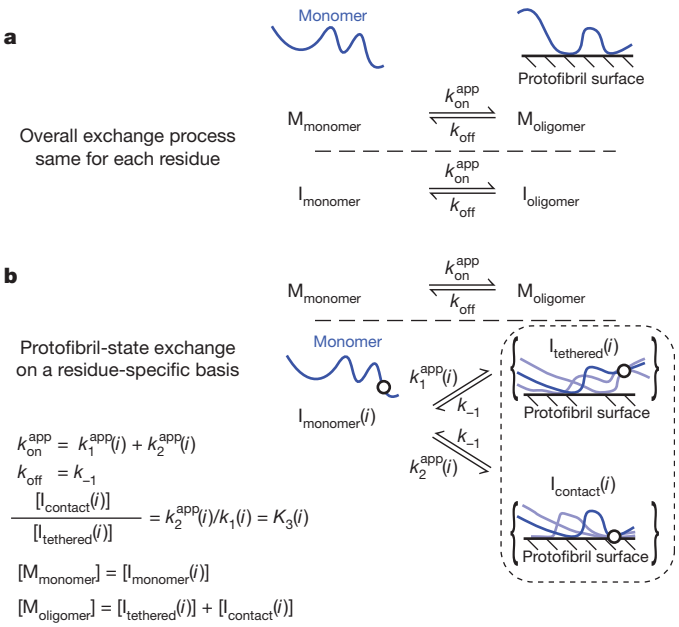


Figure 2 | Kinetic schemes for monomer exchange on the surface of Aβ protofibrils. **a**, The protofibril-bound peptide ($M_{oligomer}$) exists in only a single state. **b**, The protofibril-bound peptide exists as a large ensemble of states such that each residue can be either tethered or in direct contact with the surface of the oligomer with $K_3(i) = k_2^{app}(i)/k_1^{app}(i)$. The circle in the diagrammatic representation of the states represents a single residue that is either tethered or in direct contact and for which three possible chain configurations are shown.

©2011 Macmillan Publishers Limited. All rights reserved

with the protofibril surface ($I_{contact}$) or tethered ($I_{tethered}$) by residues in direct contact with the protofibril (Fig. 2b). The key to understanding this model lies in the observation that the variations in both ΔR_2 and ^{15}N-DEST profiles are residue specific. This variation cannot be accounted for by postulating the existence of low molecular mass oligomers because, in that case, the ΔR_2 and DEST effects would be uniform throughout the peptide sequence (as each residue would be subject to the same events and the same average relaxation behaviour when bound to the surface of the oligomers). Thus for each residue, i, there is a potentially different partitioning of the large number of conformational states into $I_{tethered}(i)$ and $I_{contact}(i)$. This partitioning is computed for a single residue at a time, the state of the other residues may be tethered, bound or inter-converting between the two, and the correlation between states at different positions is not considered. Critical to the simplicity of the model is the retention of the overall two-state association/dissociation process; the global apparent association rate constant for the conversion of monomer to the protofibril-bound ensemble, k_{on}^{app}, is given by the sum of the pseudo-first-order rate constants $k_1^{app}(i)$ and $k_2^{app}(i)$ for the conversion of a residue i from the monomeric state ($I_{monomer}$) to $I_{tethered}(i)$ and $I_{contact}(i)$, respectively; the global dissociation rate constant, k_{off}, from protofibril-bound states to monomer, is the same for $I_{tethered}(i)$ and $I_{contact}(i)$ for all residues i (that is, $k_{off} = k_{-1} = k_{-2}$). Thus, the two-state equilibrium between monomer and protofibril-bound forms is preserved, and k_{on}^{app} and k_{off} represent ensemble averaged rates. This model adds the fewest parameters to describe a system with a huge number of potential configurations (2^{40} and 2^{42} for Aβ40 and Aβ42, respectively, where each residue is represented as either tethered or in direct contact) with differing behaviour at each residue. Because there are so many combinations of states, the addition of further variation from the polydispersity of the protofibrils does not impact the interpretation of the data or its quantitative analysis in terms of the model in Fig. 2b.

A residue-specific equilibrium between the tethered and direct-contact states is described by the ratio of $I_{contact}(i)$ to $I_{tethered}(i)$, $K_3(i) = k_2^{app}(i)/k_1^{app}(i)$. Direct inter-conversion between $I_{contact}(i)$ and $I_{tethered}(i)$ is not necessary to fit the data and, hence, is not incorporated in the analysis presented in the main text (see Supplementary Table 1 and Supplementary Fig. 11 for analysis including these kinetic paths). The ^{15}N$-R_2^{tethered}$ values for the $I_{tethered}$ states are also residue-specific and describe the average dynamics of a given residue in the tethered state. In physical terms, this will depend upon the average length and flexibility of the segment separating residue i from other residues of the peptide in direct contact with the oligomer.

The model in Fig. 2b is a reduced, 'model-free' representation of the actual system, which clearly comprises a large ensemble of different states and inter-conversion networks. Although the model is phenomenological in that individual protofibril-bound states in the ensemble and the correlated behaviour of residues are not explicitly considered, the smooth variation of the fitted parameters as a function of residue is indicative of correlated behaviour at adjacent positions, as to be expected for a model that well describes physical behaviour. Further, the model does not require that a single protofibril state be present or that the members of the protofibril ensemble even have the same interfaces with bound monomers. The ability to interpret the experimentally observed residue-by-residue differences between Aβ40 and Aβ42 demonstrates the success of the model.

Least-squares nonlinear optimization of the scheme shown in Fig. 2b results in excellent fits to both the ^{15}N-DEST saturation profiles (Fig. 1b–d and Supplementary Fig. 10a–d) and the ΔR_2 data (Fig. 1e and Supplementary Fig. 10e). For Aβ40 at 270 μM, $k_{on}^{app} = 3.0 \pm 0.7\ s^{-1}$, in agreement with the observed value of $3.0\ s^{-1}$ for ΔR_2^{max} (Fig. 1e), and $k_{off} = 51 \pm 12\ s^{-1}$, in good agreement with our previous estimate of approximately $73\ s^{-1}$ obtained from a one-dimensional ^{1}H saturation experiment[15]. The values of k_{on}^{app} and k_{off} for Aβ42 at 160 μM are $2.4\ s^{-1}$ (in agreement with $2.3\ s^{-1}$ for ΔR_2^{max}) and $62\ s^{-1}$, respectively, although the standard deviations in these fitted parameters are higher owing to lower experimental signal-to-noise ($\pm 1.8\ s^{-1}$ and $\pm 46\ s^{-1}$). The values of ^{15}N$-R_2^{contact}$ are $18{,}800 \pm 700\ s^{-1}$ and $19{,}300 \pm 2000\ s^{-1}$ for Aβ40 and Aβ42, respectively, which are reasonable for the large, worm-like protofibrils seen in these samples by atomic force and electron microscopy (Supplementary Fig. 1). From the values of the association and disassociation rate constants one can deduce that the populations of transiently bound Aβ40 and Aβ42 are approximately 6% and 4%, respectively, that of free monomer.

The residue-specific partitioning, $K_3(i)$, between the direct-contact and tethered states (Fig. 3a) and the associated ^{15}N$-R_2^{tethered}$ relaxation values (Fig. 3b) provide the first characterization of the protofibril-bound forms of Aβ40 and Aβ42 at the single-residue level. Qualitatively the K_3 and ^{15}N$-R_2^{tethered}$ profiles for Aβ40 and Aβ42 are similar, with the notable exception of the ^{15}N$-R_2^{tethered}$ values for the C-terminal residues. We also note that the values of K_3 and ^{15}N$-R_2^{tethered}$ are directly correlated with the observed widths of the saturation profiles and ΔR_2, respectively (Supplementary Fig. 12), demonstrating that the model is consistent with a qualitative analysis of the data suggested directly from the observed residue-specific behaviour.

For both Aβ40 and Aβ42 the first eight N-terminal residues have similar, small values of K_3 (0.08–0.14), indicating that these residues are predominantly tethered, spending only a small fraction (≤10%) of the time in actual direct contact with the protofibril surface. Correlated with this finding, these residues retain considerable flexibility as reflected by ^{15}N$-R_2^{tethered}$ values ranging from 20 to $80\ s^{-1}$. These observations are consistent with ^{13}C linewidths in solid-state NMR spectra of Aβ40 fibrils that are indicative of conformational disorder for the N-terminal segment[10], although our data indicate a small but measurable population in direct contact with the protofibril surface.

The highest values of K_3 (>0.24) are seen for residues 16–25 and 28–37 of Aβ40 and residues 16–23 and 33–40 of Aβ42, corresponding roughly to the central and C-terminal hydrophobic regions that form intermolecular β-sheets at the core of amyloid fibrils[9,10]. Although the ratio of direct-contact to tethered states in the Aβ40 protofibril-bound ensemble is very similar for the two regions (0.34 ± 0.06 compared with 0.32 ± 0.04), the average ^{15}N$-R_2^{tethered}$ values are very different: $146 \pm 33\ s^{-1}$ for residues 16–22 falling to $61 \pm 16\ s^{-1}$ for residues 23–25, and $62 \pm 13\ s^{-1}$ for residues 28–37. This suggests that, in the tethered states, residues 28–37 of Aβ40 are considerably more flexible than residues 16–22. In terms of a physical picture this suggests that

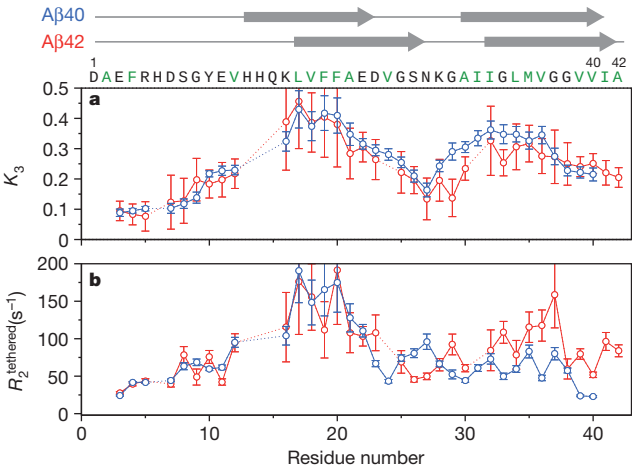


Figure 3 | Comparison of residue-specific fitted parameters describing the ensemble of protofibril-bound states. **a**, Residue-specific equilibrium constant (K_3) for the relative partitioning of direct-contact and tethered states on the surface of the protofibrils. **b**, ^{15}N$-R_2^{tethered}$ values for the tethered states. The values for Aβ40 and Aβ42 are displayed as blue and red circles, respectively. The proposed two β-strand regions in fibrils of Aβ40 and Aβ42 (refs 9 and 8, respectively) are indicated by arrows at the top of the figure. Error bars, 1 s.d.

©2011 Macmillan Publishers Limited. All rights reserved

the C-terminal region of Aβ40 may be far from the protofibril surface when other residues (such as the 16–22 region) are bound, whereas the converse is true for the 16–22 region. For Aβ42, the average value of K_3 is slightly higher in the central hydrophobic region (0.35 ± 0.06) compared with the 33–40 region (0.28 ± 0.03), but in contrast to Aβ40, the corresponding $^{15}N-R_2^{tethered}$ average values are quite similar ($133 \pm 33\ s^{-1}$ and $99 \pm 35\ s^{-1}$, respectively), suggesting that the tethered states of the C-terminal hydrophobic region are on average less mobile and closer to the protofibril surface in Aβ42 than in Aβ40.

The most striking differences between protofibril-bound Aβ40 and Aβ42 are seen at the C terminus. Intermediate K_3 values (0.21–0.23) associated with low $^{15}N-R_2^{tethered}$ values (58, 24 and 23 s^{-1} for residues 38, 39 and 40, respectively) are observed for the last three C-terminal residues of Aβ40 (Fig. 3). Thus, residues 38–40 in the ensemble of Aβ40 protofibril-bound states are more mobile and hence significantly less often buried than the central (residues 16–25) and C-terminal (residues 28–37) hydrophobic regions. This is in sharp contrast to Aβ40 fibrils, where the C-terminal residues are located at the central core of the fibre, forming tight intermolecular contacts with both neighbouring strands in the C-terminal cross-β sheet and adjacent subunits in the supramolecular structure[9]. Interestingly, the last two C-terminal residues of antibody-stabilized Aβ40 protofibrils also exhibit increased mobility as assessed by solid state NMR[19], suggesting that enhanced dynamics of the C-terminal residues of Aβ40 are at least partly conserved when the protofibril-bound states are incorporated into the protofibril core. Although the population of direct-contact states for the C-terminal residues of Aβ42 is similar to that of Aβ40 (Fig. 3a), the corresponding $^{15}N-R_2^{tethered}$ values are significantly higher for Aβ42 ($76 \pm 16\ s^{-1}$ for residues 39–42 of Aβ42 compared with $23.5 \pm 0.5\ s^{-1}$ for residues 39–40 of Aβ40; Fig. 3b). Thus, the mobility of the C-terminal residues in the tethered states of Aβ42 is hindered relative to that in Aβ40.

If dynamics of the C-terminal hydrophobic region and C-terminal residues in the protofibril-bound states, where Aβ40 and Aβ42 differ (Fig. 3), constitute the apparent free-energy barrier in the nucleation-dependent conformational change or polymerization separating disordered oligomers from fast-elongating amyloid fibrils[24], these observed differences may explain the higher rate of protofibril and fibril formation of Aβ42 relative to Aβ40 (refs 13, 14).

Further, although the population of direct-contact states for the hydrophilic region between residues 27 and 29 is reduced for both Aβ40 and Aβ42, the K_3 values are higher for Aβ40 than Aβ42 (Fig. 3a). Because this region constitutes a hairpin turn in Aβ fibrils, the lower population of direct contact states for Aβ42 may facilitate hairpin and subsequent β-sheet formation, thereby enhancing the rate of incorporation of surface-bound monomers into the protofibril core.

In summary, ^{15}N-DEST and ΔR_2 measurements probe exchange processes between monomer and protofibril-bound states of Aβ on times scale of 10–100 ms, and imprint the residue-by-residue footprint of the 'dark' protofibril-bound state on the easily observed monomer. These experiments are complementary to pulsed H/D exchange experiments that have been used to probe fibril formation and recycling on a million-fold slower timescale[25,26], as well as to spin-state selective methyl group relaxation dispersion experiments[27] that probe invisible state dynamics of supramolecular complexes on the intermediate chemical shift timescale of 50 μs to 10 ms (ref. 28). In contrast to relaxation dispersion, which requires that the visible and invisible states have different chemical shifts, the DEST and ΔR_2 experiments rely entirely on large differences in transverse relaxation rates (R_2). For systems where a relatively small particle (that is, a macromolecule whose resonances are directly observable by solution NMR) is in exchange with a state bound to a large particle (for example, an aggregate, a fibril, a large supramolecular complex or a solid support), large differences in R_2 values will be present and consequently the R_2 values in the 'dark' state can be determined by DEST. The DEST experiment requires only that the exchange rate ($k_{on}^{app} + k_{off}$) be larger than the ^{15}N-R_1 of the observed species. The Aβ monomer has a ^{15}N-R_1 of approximately $1.5\ s^{-1}$, significantly faster and hence more limiting than for a larger protein. Because k_{on}^{app} is a pseudo-first-order rate constant, the use of higher concentrations can increase the apparent on-rate, thereby extending the range of applicability. The upper limit on the value of k_{on}^{app} is approximately $100\ s^{-1}$, beyond which accurate determination of ΔR_2 is difficult. We anticipate that the DEST technique for examining dynamics at single-residue resolution of otherwise invisible 'dark' states will be applicable to many areas of current interest both in biology (for example, the interaction of unfolded proteins with the proteasome[29] and very large chaperones, the formation of low-affinity, large supramolecular complexes, and the folding-upon-binding[30] of intrinsically disordered proteins) and materials science (for example, exchange reactions on surfaces).

METHODS SUMMARY

Solutions of uniformly ^{15}N-labelled Aβ40 and Aβ42 were prepared from NaOH-treated stocks as described previously[15]. NMR samples, comprising 50 and 270 μM Aβ40 and 50 and 160 μM Aβ42 in 50 mM HEPES, pH 6.8, and 90% H_2O/10% D_2O, were prepared and maintained at 4–10 °C at all times. All NMR experiments were conducted at 10 °C on a Bruker 900 MHz spectrometer equipped with a triple resonance z-gradient cryoprobe. Fitting the DEST and ΔR_2 data to various kinetic schemes using the McConnell equations to determine the oligomer-bound R_2 values and kinetic parameters was performed as described in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 June; accepted 19 September 2011.
Published online 30 October 2011.

1. Lashuel, H. A. & Lansbury, P. T. Are amyloid diseases caused by protein aggregates that mimic bacterial pore-forming toxins? *Q. Rev. Biophys.* **39,** 167–201 (2006).
2. Walsh, D. M. & Selkoe, D. J. Aβ oligomers – a decade of discovery. *J. Neurochem.* **101,** 1172–1184 (2007).
3. Glabe, C. G. Structural classification of toxic amyloid oligomers. *J. Biol. Chem.* **283,** 29639–29643 (2008).
4. Querfurth, H. W. & LaFerla, F. M. Mechanisms of disease: Alzheimer's disease. *N. Engl. J. Med.* **362,** 329–344 (2010).
5. Ahmed, M. *et al.* Structural conversion of neurotoxic amyloid-β_{1-42} oligomers to fibrils. *Nature Struct. Mol. Biol.* **17,** 561–567 (2010).
6. Fukumoto, H. *et al.* High-molecular-weight beta-amyloid oligomers are elevated in cerebrospinal fluid of Alzheimer patients. *FASEB J.* **24,** 2716–2726 (2010).
7. Petkova, A. T. *et al.* A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl Acad. Sci. USA* **99,** 16742–16747 (2002).
8. Luhrs, T. *et al.* 3D structure of Alzheimer's amyloid-β(1–42) fibrils. *Proc. Natl Acad. Sci. USA* **102,** 17342–17347 (2005).
9. Paravastu, A. K., Leapman, R. D., Yau, W. M. & Tycko, R. Molecular structural basis for polymorphism in Alzheimer's β-amyloid fibrils. *Proc. Natl Acad. Sci. USA* **105,** 18349–18354 (2008).
10. Petkova, A. T., Yau, W. M. & Tycko, R. Experimental constraints on quaternary structure in Alzheimer's β-amyloid fibrils. *Biochemistry* **45,** 498–512 (2006).
11. Fawzi, N. L., Okabe, Y., Yap, E. H. & Head-Gordon, T. Determining the critical nucleus and mechanism of fibril elongation of the Alzheimer's $A\beta_{1-40}$ peptide. *J. Mol. Biol.* **365,** 535–550 (2007).
12. Powers, E. T. & Powers, D. L. Mechanisms of protein fibril formation: nucleated polymerization with competing off-pathway aggregation. *Biophys. J.* **94,** 379–391 (2008).
13. Jarrett, J. T., Berger, E. P. & Lansbury, P. T. Jr. The carboxy terminus of the beta amyloid protein is critical for the seeding of amyloid formation: implications for the pathogenesis of Alzheimer's disease. *Biochemistry* **32,** 4693–4697 (1993).
14. Riek, R., Guntert, P., Dobeli, H., Wipf, B. & Wuthrich, K. NMR studies in aqueous solution fail to identify significant conformational differences between the monomeric forms of two Alzheimer peptides with widely different plaque-competence, Aβ(1–40)(ox) and Aβ(1–42)(ox). *Eur. J. Biochem.* **268,** 5930–5936 (2001).
15. Fawzi, N. L., Ying, J., Torchia, D. A. & Clore, G. M. Kinetics of amyloid β monomer-to-oligomer exchange by NMR relaxation. *J. Am. Chem. Soc.* **132,** 9948–9951 (2010).
16. Teplow, D. B. *et al.* Elucidating amyloid β-protein folding and assembly: a multidisciplinary approach. *Acc. Chem. Res.* **39,** 635–645 (2006).
17. Mastrangelo, I. A. *et al.* High-resolution atomic force microscopy of soluble Aβ42 oligomers. *J. Mol. Biol.* **358,** 106–119 (2006).
18. Pimplikar, S. W. Reassessing the amyloid cascade hypothesis of Alzheimer's disease. *Int. J. Biochem. Cell Biol.* **41,** 1261–1268 (2009).
19. Scheidt, H. A., Morgado, I., Rothemund, S., Huster, D. & Fandrich, M. Solid-state NMR spectroscopic investigation of Aβ protofibrils: implication of a β-sheet remodeling upon maturation into terminal amyloid fibrils. *Angew. Chem.* **50,** 2837–2840 (2011).

©2011 Macmillan Publishers Limited. All rights reserved

20. Hou, L. M. *et al.* Solution NMR studies of the Aβ(1–40) and Aβ(1–42) peptides establish that the met35 oxidation state affects the mechanism of amyloid formation. *J. Am. Chem. Soc.* **126,** 1992–2005 (2004).
21. Yan, Y. & Wang, C. Aβ42 is more rigid than Aβ40 at the C terminus: implications for Aβ aggregation and toxicity. *J. Mol. Biol.* **364,** 853–862 (2006).
22. McConnell, H. M. Reaction rates by nuclear magnetic resonance. *J. Chem. Phys.* **28,** 430–431 (1958).
23. Helgstrand, M., Hard, T. & Allard, P. Simulations of NMR pulse sequences during equilibrium and non-equilibrium chemical exchange. *J. Biomol. NMR* **18,** 49–63 (2000).
24. Lee, J., Culyba, E. K., Powers, E. T. & Kelly, J. W. Amyloid-β forms fibrils by nucleated conformational conversion of oligomers. *Nature Chem. Biol.* **7,** 602–609 (2011).
25. Carulla, N. *et al.* Molecular recycling within amyloid fibrils. *Nature* **436,** 554–558 (2005).
26. Carulla, N., Zhou, M., Giralt, E., Robinson, C. V. & Dobson, C. M. Structure and intermolecular dynamics of aggregates populated during amyloid fibril formation studied by hydrogen/deuterium exchange. *Acc. Chem. Res.* **43,** 1072–1079 (2010).
27. Hansen, D. F., Vallurupalli, P. & Kay, L. E. Measurement of methyl group motional parameters of invisible, excited protein states by NMR spectroscopy. *J. Am. Chem. Soc.* **131,** 12745–12754 (2009).
28. Ishima, R. & Torchia, D. A. Accuracy of optimized chemical-exchange parameters derived by fitting CPMG R_2 dispersion profiles when $R_2^{0a} \neq R_2^{0b}$. *J. Biomol. NMR* **34,** 209–219 (2006).
29. Ruschak, A. M., Religa, T. L., Breuer, S., Witt, S. & Kay, L. E. The proteasome antechamber maintains substrates in an unfolded state. *Nature* **467,** 868–871 (2010).
30. Sugase, K., Dyson, H. J. & Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447,** 1021–1025 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Tycko for discussions, D. Baber, D. Garrett and M. Cai for NMR technical assistance, F. Shewmaker for performing dot blots, and W. Qiang, B. Chen and K. Thurber for assistance with atomic force microscopy and electron microscopy imaging. This work was supported by the intramural program of the National Institute of Diabetes and Digestive and Kidney Diseases/National Institutes of Health and the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (to G.M.C.).

Author Contributions All authors contributed extensively to the work described in this paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.M.C. (mariusc@mail.nih.gov).

©2011 Macmillan Publishers Limited. All rights reserved

METHODS

Description of the two-dimensional ^{15}N-DEST experiment and pulse sequence. Owing to relatively small equilibrium magnetization of ^{15}N nuclei, initial amide hydrogen ($^{1}H_N$) magnetization for the monomeric species is transferred to longitudinal ^{15}N magnetization, after which a weak ^{15}N saturation pulse is applied at a series of radio-frequency offsets (ranging from +35 to −35 kHz) for a sufficient length of time (0.9 s) to allow the monomeric peptide to sample a protofibril-bound state where ^{15}N-R_2 values are orders of magnitude larger than in the monomeric species and allow for efficient partial saturation by the weak radio-frequency field. Partial saturation is then transferred back to the monomeric state by chemical exchange, and recorded as a decrease in intensity of the ^{1}H–^{15}N correlation cross-peaks for the NMR observable monomeric species relative to that in a reference spectrum obtained without saturation (see Supplementary Fig. 8). Because initial $^{1}H_N$ to ^{15}N INEPT transfer of magnetization in the high molecular mass protofibrils is extremely inefficient owing to very fast ^{1}H and ^{15}N transverse relaxation, the observed transfer-of-saturation only arises from peptides that begin the experiment as monomers, are incorporated into protofibrils, and subsequently released within the saturation time. The pulse sequence for the two-dimensional heteronuclear single quantum coherence correlation-based ^{15}N-DEST experiment is shown in Supplementary Fig. 8. The pulse sequence is similar in spirit to that commonly used for measuring ^{15}N-R_1 longitudinal relaxation except that the variable T_1 relaxation delay is replaced by a weak, continuous-wave saturation pulse at variable radio-frequency offsets. The DEST experiment begins from the amide proton magnetization H_z of the monomeric peptide. Following the first INEPT transfer, longitudinal two spin order $2H_zN_z$ is obtained (Supplementary Fig. 8, point *a*). The subsequent refocusing INEPT provides in phase N_z magnetization (Supplementary Fig. 8, point *b*). The 90° ^{1}H pulse ϕ_2 not only flips the water signal back to the +*z*-axis, but also purges any residual anti-phase magnetization $2N_yH_z$ before the ^{15}N continuous-wave saturation pulse. The latter is applied for 0.9 s at a series of 15 different offsets (35, 28, 21, 14, 8, 4, 2, 0, −2, −4, −8, −14, −21, −28 and −35 kHz) from the ^{15}N carrier frequency (set to 118.5 and 117.6 p.p.m. and located at the centre of the monomer spectrum of Aβ40 and Aβ42, respectively) and at two radio-frequency field strengths (170 and 350 Hz). In addition, two reference spectra were recorded at each concentration without saturation. During the continuous-wave ^{15}N pulse, the magnetization (N_z) originating from the monomeric peptide is partly saturated when the peptide samples a protofibril-bound state. The latter cannot be directly observed owing to extremely rapid transverse (R_2) relaxation. This partial saturation results in a reduction of the N_z magnetization by a multiplicative factor κ (relative to that in the reference spectrum with the saturation pulse turned off) which depends on the strength and frequency of the ^{15}N saturation pulse as well as on the kinetics of the monomer-oligomer exchange process (see main text for more details), thereby providing valuable insight into the inter-conversion dynamics at atomic resolution. When the oligomer-bound peptide is released back into solution as a monomer, its reduced magnetization κN_z is detected by heteronuclear single quantum correlation. It is important to note that inversion of the phases of both the 90° ^{1}H pulse ϕ_1 and the receiver phase alternates the orientation of the N_z magnetization between the +*z* and –*z*-axes. Summation of transients with these alternating phases removes non-exponential contributions to longitudinal relaxation (for example, owing to non-zero magnetization at Boltzman equilibrium) so that loss of magnetization due to R_1 relaxation at the end of the continuous-wave saturation pulse is a simple scaling factor $e^{-R_1 t}$, and hence the R_1 dependence is effectively removed when cross-peak intensities are normalized to those in the spectrum without saturation[31]. Longitudinal cross-correlated relaxation interference between the ^{15}N–^{1}H dipolar interaction and the ^{15}N chemical shift anisotropy is eliminated by application of the 180° ^{1}H pulse train (with an inter-pulse delay of 100 ms) during the continuous-wave ^{15}N pulse. With a cryogenic probe at a ^{1}H spectrometer frequency of 900 MHz, this inter-pulse delay is sufficient to return the water signal back to the +*z*-axis through radiation damping. The de-phasing INEPT following the saturation period results in an anti-phase term $\kappa 2N_xH_z$ (Supplementary Fig. 8, point *c*). The anti-phase ^{15}N magnetization is then converted to amide proton anti-phase magnetization. The last refocusing INEPT, incorporating a WATERGATE scheme, not only re-phases the anti-phase proton magnetization to in-phase κH_x, but also suppresses the residual water signal before direct detection of the monomeric peptide amide proton magnetization.

Data acquisition and analysis. Solutions of uniformly ^{15}N-labelled Aβ40 and Aβ42 were prepared from NaOH-treated stocks as described previously[15]. NMR samples, comprising 50 and 270 μM Aβ40 and 50 and 160 μM Aβ42 in 50 mM HEPES, pH 6.8, and 90% H_2O/10% D_2O, were prepared and maintained at 4–10 °C at all times. All NMR experiments were recorded at 10 °C using a Bruker Avance-III spectrometer operating at ^{1}H frequency 900.27 MHz and equipped with a Bruker TCI *z*-axis gradient cryogenic probe. ^{15}N-R_2 data on the 50 and 270 μM Aβ40 samples and 50 and 160 μM Aβ42 samples and ^{15}N-R_1 data at 50 μM were measured using standard interleaved two-dimensional ^{1}H–^{15}N heteronuclear single quantum coherence correlation-based NMR experiments and analysed as described previously[15].

The two-dimensional ^{15}N-DEST spectra were acquired as 32 interleaved experiments (with the radio-frequency field strengths and frequencies for the continuous-wave ^{15}N pulse given above). For Aβ40, each two-dimensional experiment comprises 154* × 1,900* complex data points in the indirect ^{15}N and direct ^{1}H dimensions, respectively, with corresponding acquisition time of 77 and 175.6 ms. For Aβ42, 78* × 1,900* complex points were acquired with acquisition times of 41 and 175.6 ms. Sixteen transients for Aβ40 and 32 for Aβ42 were collected per free induction decay with a 1.7-s delay between scans. The total experimental times for Aβ40 and Aβ42 were each approximately 5 days.

The complete interleaved two-dimensional data set was first split into 32 individual two-dimensional data sets, one for each experiment at a particular ^{15}N saturation pulse power level and frequency, using a C program written in-house. Each experiment was then processed identically and analysed using NMRPipe[32]. The time-domain data in each spectrum were apodized in the indirect ^{15}N dimension with a 54°-shifted sine bell function. For the direct ^{1}H dimension, an exponential window function with a line broadening of 5 Hz for Aβ40 and 8 Hz for Aβ42 was applied, followed in the case of the Aβ40 data by a 72°-shifted squared sine bell. The additional line-broadening for the Aβ42 data was used to improve the signal-to-noise owing to the very low (approximately 35 μM) concentration of free Aβ42 monomer. Time-domain data in the ^{15}N and ^{1}H dimensions were then zero-filled to obtain a final complex 1024* × 4096* data matrix, with digital resolutions of 2.0 and 2.6 Hz, respectively. The residue-specific signal reduction factor κ as a function of frequency offset of the continuous-wave ^{15}N saturation pulse was obtained from the ratio of peak heights in the saturation experiment versus the reference experiment with the saturation pulse turned off. (The Bruker pulse sequence and acquisition parameter files for the ^{15}N-DEST experiment, as well as the processing scripts, are available on request.)

Simultaneous fitting of the DEST and ΔR_2 data to kinetic models. A homogenous form of the McConnell equations[22] describing a single spin in chemical exchange, as presented in equation (21) of ref. 23, was expanded to a three-state system. Owing to the small gyromagnetic ratio of ^{15}N nuclei and their sparse density in proteins, cross-relaxation between neighbouring ^{15}N nuclei in both monomer and protofibril-bound states is negligible, in marked contrast with the one-dimensional ^{1}H transfer-of-saturation profiles reported previously[15]. Effects on the saturation owing to cross-relaxation between ^{15}N and directly bonded ^{1}H nuclei, though predicted to be small, are removed with phase-cycling, placing the initial ^{15}N magnetization during the saturation period on the +*z* and –*z*-axes in alternating transients and measuring their intensity difference. The time dependence of the magnetization is given by:

$$\frac{d}{dt}\begin{bmatrix} E/2 \\ I_x^A \\ I_y^A \\ I_z^A \\ I_x^B \\ I_y^B \\ I_z^B \\ I_x^C \\ I_y^C \\ I_z^C \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda^A+k_1+k_2 & \Omega^A & -\omega_y & -k_{-1} & 0 & 0 & -k_{-2} & 0 & 0 \\ 0 & -\Omega^A & \lambda^A+k_1+k_2 & \omega_x & 0 & -k_{-1} & 0 & 0 & -k_{-2} & 0 \\ -2\Theta^A & \omega_y & -\omega_x & \rho^A+k_1+k_2 & 0 & 0 & -k_{-1} & 0 & 0 & -k_{-2} \\ 0 & -k_1 & 0 & 0 & \lambda^B+k_{-1}+k_3 & \Omega^B & -\omega_y & -k_{-3} & 0 & 0 \\ 0 & 0 & -k_1 & 0 & -\Omega^B & \lambda^B+k_{-1}+k_3 & \omega_x & 0 & -k_{-3} & 0 \\ -2\Theta^B & 0 & 0 & -k_1 & \omega_y & -\omega_x & \rho^B+k_{-1}+k_3 & 0 & 0 & -k_{-3} \\ 0 & -k_2 & 0 & 0 & -k_3 & 0 & 0 & \lambda^C+k_{-2}+k_{-3} & \Omega^C & -\omega_y \\ 0 & 0 & -k_2 & 0 & 0 & -k_3 & 0 & -\Omega^C & \lambda^C+k_{-2}+k_{-3} & \omega_x \\ -2\Theta^C & 0 & 0 & -k_2 & 0 & 0 & -k_3 & \omega_y & -\omega_x & \rho^C+k_{-2}+k_{-3} \end{bmatrix} \times \begin{bmatrix} E/2 \\ I_x^A \\ I_y^A \\ I_z^A \\ I_x^B \\ I_y^B \\ I_z^B \\ I_x^C \\ I_y^C \\ I_z^C \end{bmatrix} \quad (1)$$

where *I* represents the magnetizations of the ^{15}N spins in the rotating frame in the free monomer, and in the tethered and direct-contact protofibril-bound states (denoted by the superscripts *A*, *B* and *C*, respectively). The relaxation rates for transverse and longitudinal magnetization are λ and ρ, respectively; Ω is the resonance offset frequency calculated as the difference between the peak resonance frequency (accounting for peak folding in the case of Aβ42 Ala 42) and

©2011 Macmillan Publishers Limited. All rights reserved

the ^{15}N carrier frequency + continuous-wave pulse offset, and ω the radio-frequency field strength of the continuous-wave saturation pulse applied along the appropriate axis; E is unity; and Θ is related to the equilibrium magnetization as described by Helgstrand *et al.*[23]. The rate constants are as follows (see Fig. 2b): k_1 and k_2 are the apparent first-order rate constants for the forward reaction from the monomer to the tethered and direct-contact protofibril-bound states, respectively; k_{-1} and k_{-2} are the corresponding dissociation rate constants and in this instance $k_{-1} = k_{-2}$ (see main text); k_3 and k_{-3} (not shown in Fig. 2b) are the rate constants for the conversion of tethered to direct-contact states, and vice versa, respectively. Detailed balance is ensured from the relationship $k_{-3} = k_3k_1/k_2$ (with $k_{-1} = k_{-2}$). The numerical solution for I_z^A, the experimentally observed magnetization in the DEST experiment, after a given saturation time as a function of frequency offset was calculated using the matrix exponential function in the program Matlab. The solution is calculated for both initial ^{15}N magnetization during the saturation period on the $+z$ and $-z$-axes, I_{+z} and I_{-z}, and the difference, $I_{+z} - I_{-z}$, is then computed at each frequency offset and radio-frequency field strength, normalized to the solution without saturation, and compared to the normalized experimentally observed intensity.

ΔR_2 values, for a given set of kinetic and relaxation rate constants, were calculated by subtracting the monomer ^{15}N-R_2 values for each residue, measured experimentally on the monomeric 50 μM samples, from the calculated R_2 values. The calculated R_2 value for a system in equilibrium exchange is computed assuming a single exponential decay:

$$R_2^{\text{calc}} = -\ln\left(\frac{I_x^A(\tau)}{I_x^A(0)}\right)/\tau \qquad (2)$$

where the transverse magnetization, I_x^A, is computed at time $\tau = 0.2$ s and time 0. The choice of 0.2 s for τ was to match to one of the longer delays used in the experimental ^{15}N-R_2 measurements. Simulated delays longer than approximately 20 ms result in the same value of R_2^{calc}, validating the assumption of single exponential decay.

^{15}N relaxation rates in the protofibril-bound states, and kinetic rate constants connecting monomer and protofibril-bound states in all schemes (cf. Fig. 2), were fitted by simultaneous nonlinear least squares minimization of the differences between the observed and calculated (equation (1)) DEST profiles (at both applied radio-frequency field strengths for the continuous-wave ^{15}N saturation pulse) and between the observed and calculated (equation (2)) ΔR_2 values.

31. Sklenar, V., Torchia, D. & Bax, A. Measurement of ^{13}C longitudinal relaxation using ^{1}H detection. *J. Magn. Reson.* **73,** 375–379 (1987).
32. Delaglio, F. *et al.* NmrPipe – a multidimensional spectral processing system based on Unix pipes. *J. Biomol. NMR* **6,** 277–293 (1995).

©2011 **Macmillan Publishers Limited. All rights reserved**

LETTER

doi:10.1038/nature10557

Aspartate 112 is the selectivity filter of the human voltage–gated proton channel

Boris Musset[1]*, Susan M. E. Smith[2]*, Sindhu Rajan[3], Deri Morgan[1], Vladimir V. Cherny[1] & Thomas E. DeCoursey[1]

The ion selectivity of pumps and channels is central to their ability to perform a multitude of functions. Here we investigate the mechanism of the extraordinary selectivity of the human voltage-gated proton channel[1], H_V1 (also known as HVCN1). This selectivity is essential to its ability to regulate reactive oxygen species production by leukocytes[2–4], histamine secretion by basophils[5], sperm capacitation[6], and airway pH[7]. The most selective ion channel known, H_V1 shows no detectable permeability to other ions[1]. Opposing classes of selectivity mechanisms postulate that (1) a titratable amino acid residue in the permeation pathway imparts proton selectivity[1,8–11], or (2) water molecules 'frozen' in a narrow pore conduct protons while excluding other ions[12]. Here we identify aspartate 112 as a crucial component of the selectivity filter of H_V1. When a neutral amino acid replaced Asp 112, the mutant channel lost proton specificity and became anion-selective or did not conduct. Only the glutamate mutant remained proton-specific. Mutation of the nearby Asp 185 did not impair proton selectivity, indicating that Asp 112 has a unique role. Although histidine shuttles protons in other proteins, when histidine or lysine replaced Asp 112, the mutant channel was still anion-permeable. Evidently, the proton specificity of H_V1 requires an acidic group at the selectivity filter.

Voltage-gated proton channels are considered specific (perfectly selective) for protons, because no evidence exists for permeation of anything but H^+. Specificity, combined with a large deuterium isotope effect[9] and extraordinarily strong temperature dependence of conduction[10] suggests a permeation pathway more complex than a simple water wire, as exists in gramicidin[13]. All proton conduction seems consistent with a hydrogen-bonded chain (HBC) mechanism[14]; a HBC including a titratable group could explain several unique properties of H_V1 (ref. 1), especially proton selectivity[14]. Yet in a recent study, mutation of each titratable amino acid in all four transmembrane helices of H_V1 failed to abolish conduction[12]. Thus, the mechanism producing proton selectivity remained unknown.

We noticed that a human gene, *C15orf27* (of unknown function), contains a predicted voltage sensor domain (VSD) that shares 25% sequence identity and 52% similarity (http://www.ebi.ac.uk/Tools/emboss/align/needle) with the VSD of H_V1, and includes three Arg residues in the S4 transmembrane helix that are conserved among all known H_V1 homologues. Phylogenetic analysis of VSD sequences (Supplementary Fig. 1) reveals that a group comprising H_V1, C15orf27 and voltage-sensitive phosphatase (VSP) sequences separated early from the two phylogenetically distinct groups of depolarization activated VSDs described previously (K_V channels and Na_V/Ca_V channels), supporting the modular evolution of VSD-containing proteins[15]. Furthermore, H_V1 VSDs occupy a discrete lineage, distinct from those of VSP and C15orf27 orthologues.

When we cloned the *C15orf27* gene and expressed the product in HEK-293 or COS-7 cells, the green fluorescent protein (GFP)-tagged protein localized at the plasma membrane (Supplementary Fig. 2), but we detected no currents beyond those in non-transfected cells. We reasoned that substitutions based on sequence elements that differ between H_V1 and C15orf27 should be structurally tolerated while revealing residues responsible for proton conduction. We therefore mutated residues that are perfectly conserved in 21 H_V1 family members and differ between C15orf27 and H_V1. We replaced five candidate residues in H_V1 (D112, D185, N214, G215 and S219) (Fig. 1a, b) with the corresponding residue in C15orf27. Four mutants exhibited large currents under whole-cell voltage clamp (Fig. 1d). The reversal (zero current) potential, V_{rev}, measured at several pH_o and pH_i (external and internal pH, respectively), was close to the Nernst potential for protons, E_H (Fig. 1c), demonstrating proton selectivity. D112V mutants localized to the plasma membrane (Supplementary Fig. 3), but showed no convincing current (Fig. 1d). Some D112V-transfected HEK-293 or COS-7 cells (and non-transfected cells) had small native proton currents. H140A/H193A double mutants[16,17], in which the two Zn^{2+}-binding His residues are neutralized, resemble wild type, with similar ΔpH-dependent gating[12], and V_{rev} near E_H (Supplementary Fig. 4). We expressed mutants in this Zn^{2+}-insensitive background (D112X/A/A) to distinguish their currents from native currents that are abolished by 100 μM Zn^{2+} at pH_o 7.0. We tentatively concluded that Asp 112 is crucial to proton conduction.

The absence of detectable currents in D112V led us to make other D112X substitutions. These mutants (Fig. 2a) showed slowly activating outward currents upon depolarization that resembled H_V1 currents. As reported previously[12], Asp 112 mutation had little effect on the ΔpH dependence of gating. The proton conductance–voltage (g_H–V) relationship of all D112X mutants shifted roughly −60 mV when pH_o increased from 5.5 to 7.0 (Supplementary Fig. 5), as in wild-type channels[1,8,18]. Mutation of Asp 112 did influence channel opening and closing kinetics (Supplementary Table 2).

Measurements of V_{rev} in Asp 112 mutants showed a marked departure from wild-type H_V1 properties. At symmetrical pH 5.5, V_{rev} was near 0 mV (not shown). At pH_o 7.0, pH_i 5.5 (Fig. 2a, column 3), wild-type channels reversed near E_H (−87 mV), indicating proton selectivity. But for all mutants except D112E, V_{rev} was substantially positive to E_H (Fig. 2b), ranging from −58 mV (D112H) to −13 mV (D112N). Substitutions at Asp 112 eliminated the proton specificity that distinguishes H_V1 from all other ion channels[1]. A previous study described currents in D112A and D112N mutants[12], but did not report V_{rev}.

We expected that loss of proton selectivity would result in nonselective permeation of cations. Surprisingly, V_{rev} did not change detectably when Na^+, K^+, *N*-methyl-D-glucamine$^+$ or TEA^+ (tetraethylammonium$^+$) replaced TMA^+ (tetramethylammonium$^+$) (Supplementary Table 4). To test anion against cation selectivity, we adopted the classical tactic of replacing a fraction of the bath solution with isotonic sucrose[19]. The Nernst equation predicts that dilution of all extracellular ions except H^+ and OH^- (leaving internal ion concentrations unchanged) will shift V_{rev} negatively for a cation-selective channel, but positively for an anion-selective one. Despite the tenfold reduction of buffer concentration, direct measurement confirmed that pH remained constant.

[1]Department of Molecular Biophysics & Physiology, Rush University Medical Center, Chicago, Illinois 60612, USA. [2]Department of Pathology, Emory School of Medicine, Atlanta, Georgia 30322, USA. [3]Department of Medicine, University of Chicago, Chicago, Illinois 60637, USA.
*These authors contributed equally to this work.

©2011 Macmillan Publishers Limited. All rights reserved

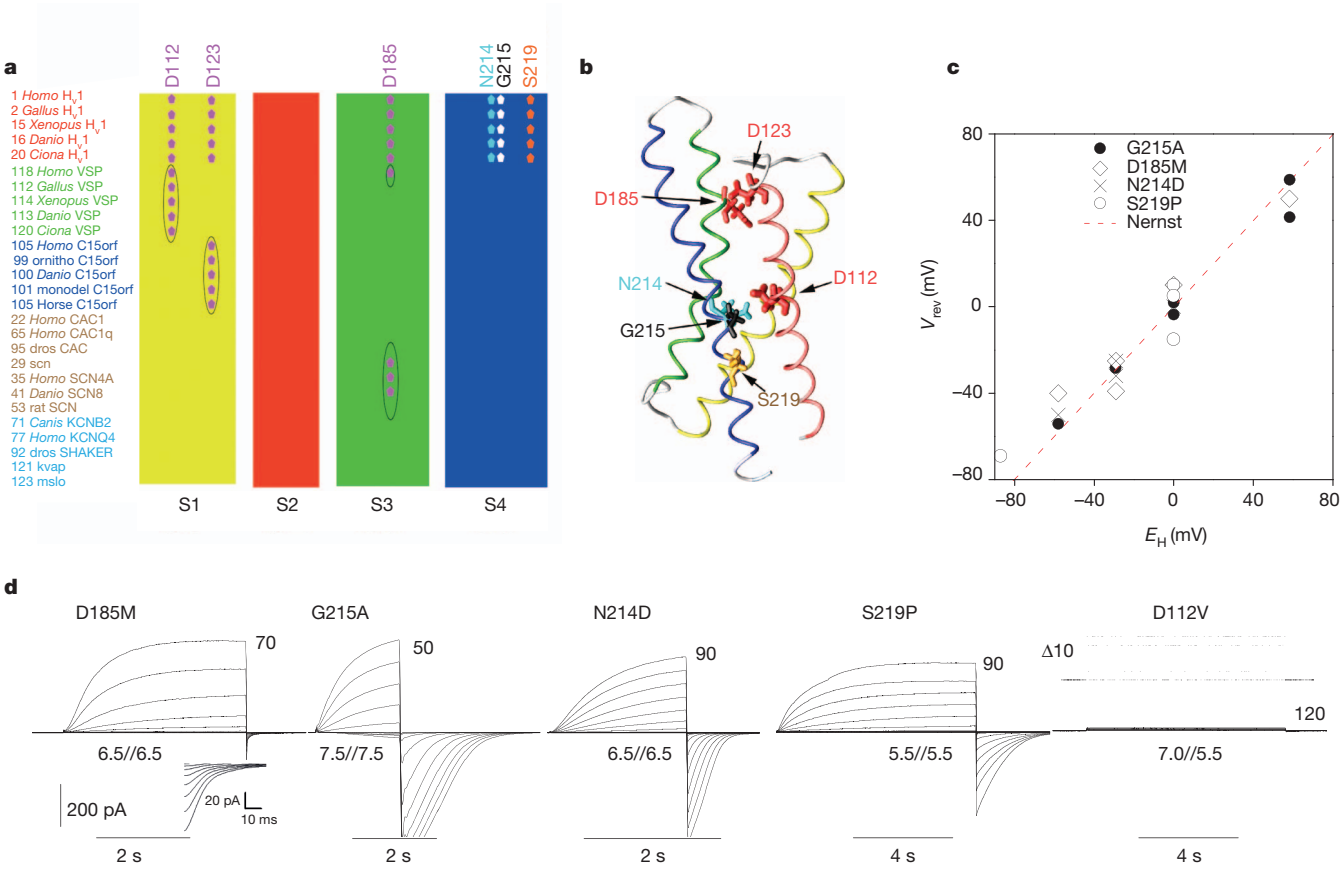


Figure 1 | Identification of five key amino acids that differ in H_V1 and C15orf27, and the currents generated in a heterologous expression system by H_V1 mutants in which H_V1 residues were replaced by the corresponding amino acid in the non-conducting C15orf27. **a**, Representative subset of multiple sequence alignment of 122 VSDs, only transmembrane helices are shown. Gene families include H_V1, voltage-sensitive phosphatases, C15orf27, Ca^{2+} and Na^+ channels, and K^+ channels (see Supplementary Fig. 1). **b**, Location of the key amino acids in the open H_V1 channel VSD viewed from the side (membrane), based on a homology model[17]. **c**, V_{rev} in the four conducting mutants is near E_H (dashed line), indicating proton selectivity. V_{rev} was measured using tail currents; in G215A V_{rev} was positive to threshold and was observed directly. **d**, Voltage-clamp current families in cells expressing H_V1 mutants. Depolarizing pulses were applied in 10-mV increments from a holding voltage, $V_{hold} = -40$ mV (D185M, D112V), -60 mV (G215A, S219P), or -90 mV (N214D), with the most positive pulse labelled. After membrane repolarization, an inward 'tail current' is seen as channels close (see inset for D185M); pH is given as $pH_o//pH_i$. D112V showed no clear current.

Figure 3 illustrates determination of V_{rev} from tail currents in a D112H-transfected cell at pH 5.5//5.5 ($pH_o//pH_i$) in methanesulphonate$^-$ ($CH_3SO_3^-$; Fig. 3a) or Cl^- solution (Fig. 3c), and after 90% reduction of external ionic strength (Figs 3b, d). Surprisingly, for all Asp 112 mutants, sucrose shifted V_{rev} positively (Supplementary Fig. 6), indicating anion selectivity both in $CH_3SO_3^-$ (Fig. 3e) and Cl^- solutions (Fig. 3f). For H_V1 and D112E, V_{rev} did not change, reaffirming their proton specificity. Neutralization of a single Asp residue converts a proton channel into a predominantly anion selective channel. Thus, Asp 112 mediates charge selectivity as well as proton selectivity.

To confirm anion permeability of Asp 112 mutants, we replaced the main external anion, $CH_3SO_3^-$, with Cl^-. Consistent with previous studies[1], V_{rev} in H_V1 was unchanged. As shown in D112H (Fig. 3a, c), V_{rev} shifted negatively in Cl^- solutions in all mutants (except D112E), indicating that Cl^- is more permeant than the larger $CH_3SO_3^-$ anion (Fig. 3g). That all conducting non-acidic mutants showed Cl^- permeability indicates that Asp 112 mediates not only proton selectivity, but also charge selectivity. Currents were smaller than wild type in cells expressing some mutants (Supplementary Fig. 7), suggesting a smaller unitary conductance. Evidently, these channels conduct anions, but not very well.

Although the mutant channels have diminished selectivity, V_{rev} did shift negatively when pH_o increased from 5.5 to 7.0 (Fig. 2b). Because these solutions differ mainly in buffer species and concentrations of H^+ and OH^-, Asp 112 mutants must have significant permeability to H^+ and/or OH^-. The Goldman–Hodgkin–Katz equation shows how V_{rev} depends on ion concentrations:

$$V_{rev} = \frac{RT}{zF}\log\frac{P_{Cl^-}[Cl^-]_i + P_{CH_3SO_3^-}[CH_3SO_3^-]_i + P_{OH^-}[OH^-]_i + P_{H^+}[H^+]_o}{P_{Cl^-}[Cl^-]_o + P_{CH_3SO_3^-}[CH_3SO_3^-]_o + P_{OH^-}[OH^-]_o + P_{H^+}[H^+]_i},$$

where R is the gas constant, T the absolute temperature (Kelvin), z the ionic valence ($= 1$), F is Faraday's constant and P is permeability.

Ions with greater permeability dominate V_{rev}. Permeation of H^+ and OH^- are difficult to distinguish because they have the same Nernst potential[1]. The data can be interpreted assuming permeation of either (Supplementary Table 3), but the anion selectivity of Asp 112 mutants and the pH dependence of sucrose effects (Figs 3e, f) support OH^- permeation. The relative permeability of conducting Asp 112 mutants was OH^- (or H^+) $> Cl^- > CH_3SO_3^-$.

Although Asp 112 is essential to selectivity, other acidic groups might participate. We mutated Asp 185, located in the presumed conduction pore (Fig. 1b)[12,17]. However, like D185M (Fig. 1b), D185V, D185A and D185N remained proton-selective (Supplementary Fig. 8). As evidence against additive effects, the double mutant D112N/D185M did not differ from D112N (Supplementary Fig. 8).

©2011 Macmillan Publishers Limited. All rights reserved

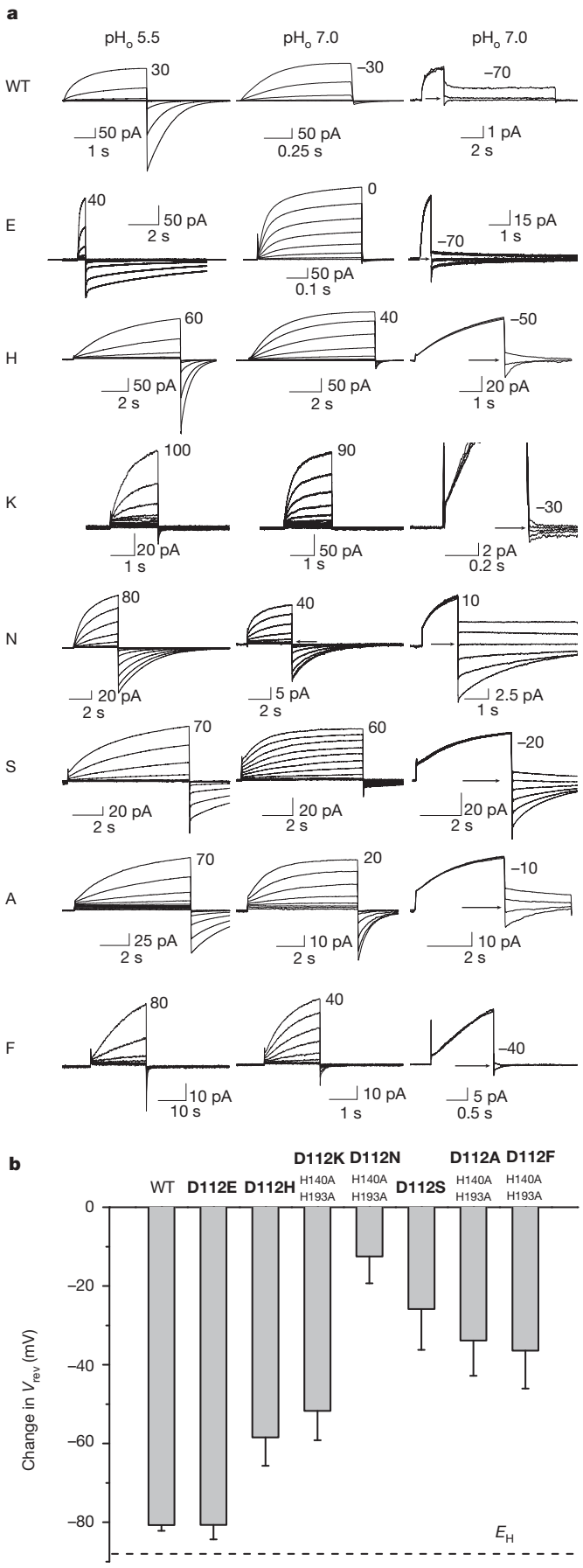


Figure 2 | Currents in Asp 112 mutants resemble proton currents, but are not. **a**, Currents generated by wild type (WT), D112E, D112H, D112K/A/A, D112N/A/A, D112S, D112A/A/A and D112F/A/A in COS-7 cells (pH$_i$ 5.5) at pH$_o$ 5.5 (column 1) or 7.0 (column 2), during families of pulses in 10 mV increments up to indicated voltages. Tail currents at pH$_o$ 7.0 (column 3) reveal that V_{rev} deviates from E_H, indicating loss of proton selectivity. At pH$_o$ 5.5 V_{hold} was −40 mV (−60 mV for WT). At pH$_o$ 7.0 V_{hold} was −40 mV (K, N, S), −50 mV (F), −60 mV (H, A), −80 mV (E), or −90 mV (WT); V_{pre} was −65 mV (WT), −40 mV (E), −10 mV (H), +50 mV (K), +40 mV (N, S), or +20 mV (A, F). V_{rev} (arrows) was determined from the amplitude and direction of tail current decay. For D112N, V_{rev} was above $V_{threshold}$ and was evident during pulse families. **b**, Shift in V_{rev} when the $TMACH_3SO_3$ bath solution was changed from pH 5.5 to 7.0. There is no difference between WT and D112E, but the shift in all other mutants is smaller than WT ($P < 0.001$, by one-way ANOVA followed by Tukey's test, $n = 7, 4, 9, 8, 6, 7, 9$ and 4). Error bars in **b** are s.e. Dashed line shows E_H.

Consistent with earlier predictions that a titratable amino acid provides the selectivity filter of H_V1 (refs 1, 8–11), only channels with acidic residues (Glu or Asp) at position 112 manifested proton specificity. Asp 112 lies at the constriction of the presumed pore (Fig. 1b), a logical location for a selectivity filter, and just external to the postulated gating charge transfer centre[20]. Our original prediction envisioned selectivity arising from protonation/deprotonation of a residue during conduction, but other mechanisms are possible. For example, proton selectivity of the influenza A M_2 viral proton channel has been explained by (1) immobilized water[21], (2) successive proton transfer and release by His 37 (refs 22, 23) and (3) delocalization of the proton among His 37 and nearby water molecules[24].

The Cl^- permeability of D112H was completely unexpected, given strong precedents for His imparting proton selectivity to channels. Histidine shuttles protons in K^+ or Na^+ channel VSDs with Arg→His mutations[25–27], in carbonic anhydrase[28] and in M_2 channels[22,23]. However, these molecules are not proton-specific[27,29]. Evidently, His shuttles protons, but does not guarantee proton selectivity. In H_V1, Asp 112 (or Glu 112 in D112E) excludes anions, resulting in proton-specific conduction. When protonated, Glu and Asp are neutral whereas His is cationic, which may explain why D112H fails to exclude anions.

The anion selectivity of neutral Asp 112 mutants indicates that electrostatic forces due to the charge distribution in the rest of the channel deter cation permeation, and that the cation selectivity of the wild-type channel is due to the anionic charge of Asp 112. Asp 185 does not participate directly in selectivity (Supplementary Fig. 8). VSP family members possess the equivalent of Asp 112 (Fig. 1a), yet conduct no current[30], illustrating that Asp 112 requires a specific microenvironment to achieve selectivity. Although permeation of Cl^- and $CH_3SO_3^-$ suggests a wide pore in D112X mutants, local geometry might differ in wild-type channels due to the presence of anionic Asp 112.

Regulation of voltage gating by ΔpH is distinct from permeation. Pathognomonic of H_V1 is a strict correlation between the g_H–V relationship and V_{rev}, in which $V_{threshold}$ shifts 40 mV per unit change in ΔpH (ref. 8). The ΔpH dependence persisted in mutants with shifted g_H–V relationships[12]. Here we show uncoupling of V_{rev} and voltage gating. Asp 112 mutants retained normal ΔpH dependence (Supplementary Fig. 5), despite the dissociation of V_{rev} from ΔpH (Fig. 2 and Supplementary Fig. 9). This uncoupling of pH control of gating from permeation speaks against any mechanism that invokes regulation by local proton concentration in the vicinity of S4 Arg residues[12].

In summary, Asp 112 is a critical component of the selectivity filter of H_V1, crucial to both proton selectivity and charge selectivity. That D112E was proton-selective, but D112H conducted anions indicates that this proton channel requires an acid at the selectivity filter. That neutralization of nearby Asp 185 did not affect selectivity suggests that Asp 112 has a unique role.

©2011 Macmillan Publishers Limited. All rights reserved

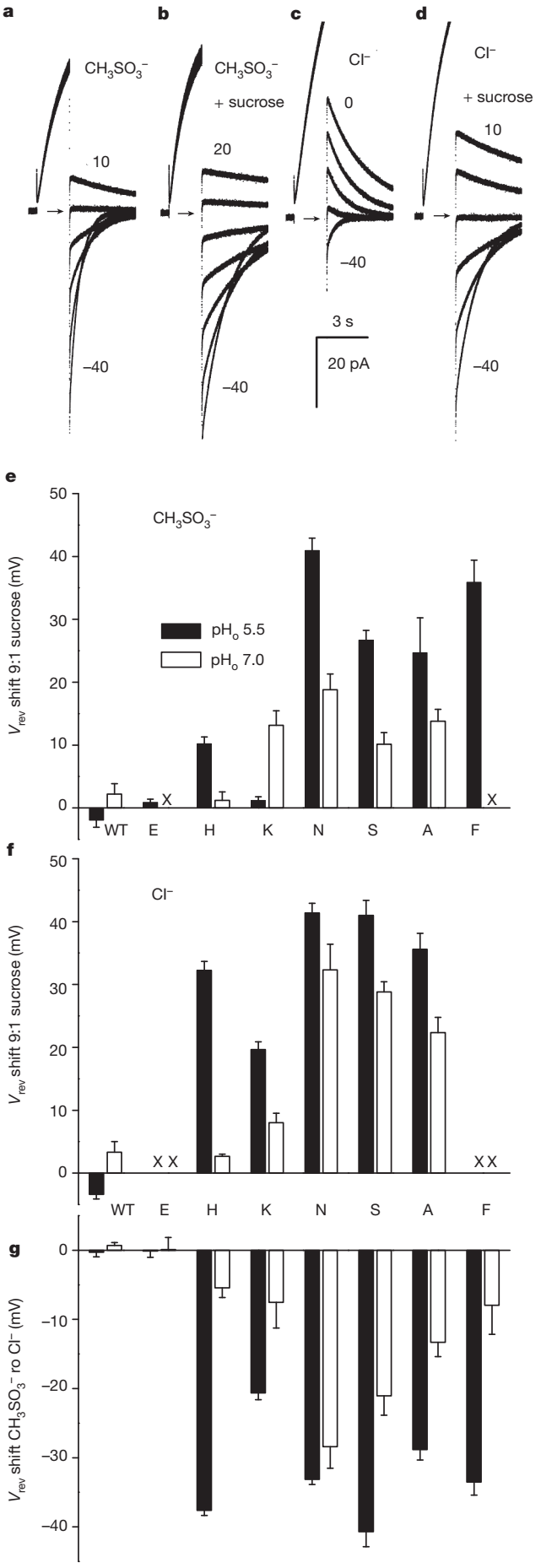


Figure 3 | Dilution of ionic strength by 90% with isotonic sucrose shifted V_{rev} positively, indicating that most Asp 112 mutants are anion-selective. **a**, Measurement of V_{rev} by tail currents in a cell transfected with D112H at pH 5.5//5.5, and **b**, after sucrose. **c**, V_{rev} in the same cell in pH 5.5 Cl^- solution, and **d**, after sucrose. Arrows indicate zero current. $V_{hold} = -40$ mV, $V_{pre} = +60$ mV. **e**, Mean shifts of V_{rev} with decreasing ionic strength in $CH_3SO_3^-$ solutions or **f**, in Cl^- solutions. Each value was determined in 3–6 cells. X = not done. **g**, Shifts of V_{rev} when $CH_3SO_3^-$ was replaced by Cl^-. Values for WT and D112E do not differ significantly from 0 mV. For all anion-selective mutants except D112N, the difference between shifts at pH 5.5 and 7.0 was significant ($P < 0.001$, one-way ANOVA followed by Tukey's test; $n = 3$–8). Error bars in **e**–**g** are s.e.

METHODS SUMMARY

The pipette solution (also used externally) contained (in mM) 130 $TMACH_3SO_3$, 2 $MgCl_2$, 2 EGTA, 80 MES (2-(*N*-morpholino)ethanesulphonic acid), titrated to pH 5.5 with ~20 TMAOH. In the pH 5.5 TMACl solution, TMACl replaced $TMACH_3SO_3$. Bath solutions at pH 7.0 had (mM) 90 $TMACH_3SO_3$ or TMACl, 3 $CaCl_2$, 1 EGTA, 100 BES and 36–40 TMAOH. For experiments with Zn^{2+}, solutions contained PIPES without EGTA. Experiments were done at 20–25 °C. Currents are shown without leak correction. V_{rev} data were corrected for liquid junctions potentials measured in each solution[19].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 April; accepted 13 September 2011.
Published online 23 October 2011.

1. DeCoursey, T. E. Voltage-gated proton channels and other proton transfer pathways. *Physiol. Rev.* **83,** 475–579 (2003).
2. Capasso, M. *et al.* HVCN1 modulates BCR signal strength via regulation of BCR-dependent generation of reactive oxygen species. *Nature Immunol.* **11,** 265–272 (2010).
3. DeCoursey, T. E., Morgan, D. & Cherny, V. V. The voltage dependence of NADPH oxidase reveals why phagocytes need proton channels. *Nature* **422,** 531–534 (2003).
4. Henderson, L. M., Chappell, J. B. & Jones, O. T. G. The superoxide-generating NADPH oxidase of human neutrophils is electrogenic and associated with an H^+ channel. *Biochem. J.* **246,** 325–329 (1987).
5. Musset, B. *et al.* A pH-stabilizing role of voltage-gated proton channels in IgE-mediated activation of human basophils. *Proc. Natl Acad. Sci. USA* **105,** 11020–11025 (2008).
6. Lishko, P. V., Botchkina, I. L., Fedorenko, A. & Kirichok, Y. Acid extrusion from human spermatozoa is mediated by flagellar voltage-gated proton channel. *Cell* **140,** 327–337 (2010).
7. Iovannisci, D., Illek, B. & Fischer, H. Function of the HVCN1 proton channel in airway epithelia and a naturally occurring mutation, M91T. *J. Gen. Physiol.* **136,** 35–46 (2010).
8. Cherny, V. V., Markin, V. S. & DeCoursey, T. E. The voltage-activated hydrogen ion conductance in rat alveolar epithelial cells is determined by the pH gradient. *J. Gen. Physiol.* **105,** 861–896 (1995).
9. DeCoursey, T. E. & Cherny, V. V. Deuterium isotope effects on permeation and gating of proton channels in rat alveolar epithelium. *J. Gen. Physiol.* **109,** 415–434 (1997).
10. DeCoursey, T. E. & Cherny, V. V. Temperature dependence of voltage-gated H^+ currents in human neutrophils, rat alveolar epithelial cells, and mammalian phagocytes. *J. Gen. Physiol.* **112,** 503–522 (1998).
11. DeCoursey, T. E. & Cherny, V. V. Voltage-activated hydrogen ion currents. *J. Membr. Biol.* **141,** 203–223 (1994).
12. Ramsey, I. S. *et al.* An aqueous H^+ permeation pathway in the voltage-gated proton channel Hv1. *Nature Struct. Mol. Biol.* **17,** 869–875 (2010).
13. Levitt, D. G., Elias, S. R. & Hautman, J. M. Number of water molecules coupled to the transport of sodium, potassium and hydrogen ions via gramicidin, nonactin or valinomycin. *Biochim. Biophys. Acta* **512,** 436–451 (1978).
14. Nagle, J. F. & Morowitz, H. J. Molecular mechanisms for proton transport in membranes. *Proc. Natl Acad. Sci. USA* **75,** 298–302 (1978).
15. Nelson, R. D., Kuan, G., Saier, M. H. Jr & Montal, M. Modular assembly of voltage-gated channel proteins: a sequence analysis and phylogenetic study. *J. Mol. Microbiol. Biotechnol.* **1,** 281–287 (1999).
16. Ramsey, I. S., Moran, M. M., Chong, J. A. & Clapham, D. E. A voltage-gated proton-selective channel lacking the pore domain. *Nature* **440,** 1213–1216 (2006).
17. Musset, B. *et al.* Zinc inhibition of monomeric and dimeric proton channels suggests cooperative gating. *J. Physiol. (Lond.)* **588,** 1435–1449 (2010).
18. Musset, B. *et al.* Detailed comparison of expressed and native voltage-gated proton channel currents. *J. Physiol. (Lond.)* **586,** 2477–2486 (2008).
19. Barry, P. H. The reliability of relative anion-cation permeabilities deduced from reversal (dilution) potential measurements in ion channel studies. *Cell Biochem. Biophys.* **46,** 143–154 (2006).
20. Tao, X., Lee, A., Limapichat, W., Dougherty, D. A. & MacKinnon, R. A gating charge transfer center in voltage sensors. *Science* **328,** 67–73 (2010).

©2011 Macmillan Publishers Limited. All rights reserved

21. Sansom, M. S. P., Kerr, I. D., Smith, G. R. & Son, H. S. The influenza A virus M2 channel: a molecular modeling and simulation study. *Virology* **233,** 163–173 (1997).
22. Hu, F., Luo, W. & Hong, M. Mechanisms of proton conduction and gating in influenza M2 proton channels from solid-state NMR. *Science* **330,** 505–508 (2010).
23. Venkataraman, P., Lamb, R. A. & Pinto, L. H. Chemical rescue of histidine selectivity filter mutants of the M2 ion channel of influenza A virus. *J. Biol. Chem.* **280,** 21463–21472 (2005).
24. Acharya, R. *et al.* Structure and mechanism of proton transport through the transmembrane tetrameric M2 protein bundle of the influenza A virus. *Proc. Natl Acad. Sci. USA* **107,** 15075–15080 (2010).
25. Starace, D. M. & Bezanilla, F. Histidine scanning mutagenesis of basic residues of the S4 segment of the *Shaker* K^+ channel. *J. Gen. Physiol.* **117,** 469–490 (2001).
26. Starace, D. M. & Bezanilla, F. A proton pore in a potassium channel voltage sensor reveals a focused electric field. *Nature* **427,** 548–553 (2004).
27. Sokolov, S., Scheuer, T. & Catterall, W. A. Ion permeation and block of the gating pore in the voltage sensor of $Na_V1.4$ channels with hypokalemic periodic paralysis mutations. *J. Gen. Physiol.* **136,** 225–236 (2010).
28. Tu, C. K., Silverman, D. N., Forsman, C., Jonsson, B. H. & Lindskog, S. Role of histidine 64 in the catalytic mechanism of human carbonic anhydrase II studied with a site-specific mutant. *Biochemistry* **28,** 7913–7918 (1989).
29. Leiding, T., Wang, J., Martinsson, J., DeGrado, W. F. & Årsköld, S. P. Proton and cation transport activity of the M2 proton channel from influenza A virus. *Proc. Natl Acad. Sci. USA* **107,** 15409–15414 (2010).
30. Murata, Y., Iwasaki, H., Sasaki, M., Inaba, K. & Okamura, Y. Phosphoinositide phosphatase activity coupled to an intrinsic voltage sensor. *Nature* **435,** 1239–1243 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. H. Barry, D. Gillespie, V. S. Markin, J. F. Nagle, R. Pomès, D. Silverman and V. Sokolov for discussions or comments on the manuscript. Supported by NSF grant MCB-0943362 (S.M.E.S. and T.E.D.) and NIH grant GM087507 (T.E.D.). The content is solely the responsibility of the authors and does not necessarily represent the views of the National Institute of General Medical Sciences or the National Institutes of Health.

Author Contributions S.R. identified the similarity of C15orf27 to H_V1 and cloned the *C15orf27* gene; S.M.E.S. conceived the strategic approach based on molecular model, sequence and phylogenetic analysis; S.R. and S.M.E.S. created mutants; T.E.D., B.M. and V.V.C. designed experiments; B.M., D.M. and V.V.C. recorded, analysed and interpreted data; T.E.D. wrote the manuscript; all authors read and approved the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.E.D. (tdecours@rush.edu).

©2011 Macmillan Publishers Limited. All rights reserved

METHODS

Exhaustive searches to identify H_V1 homologues were performed using protein BLAST and PSI-BLAST. A sample of VSDs from K^+, Na^+ and Ca^{2+} channels (that open with depolarization like H_V1, and in addition one that opens with hyperpolarization), along with putative H_V1, VSP and C15orf27 homologues were chosen. For cation channels, we sampled from the range of subfamilies, from the VSD repeats within Na^+ and Ca^{2+} channels, and from the range of species. VSD sequences, including crystallized K^+ channels (PDB accessions 1ORS, 2R9R and 2A79), were aligned using PromalS3D[31], which incorporates structural information, allowing high-confidence identification of VSD boundaries. Sequences were trimmed to the VSD, realigned with PromalS3D, and the resulting alignment was analysed with PhyML (maximum likelihood)[32] and Protpars (maximum parsimony)[33] at the Mobyle portal[34]. Trees were visualized with TreeDyn[35] and iTOL[36]. Parsimony (not shown) and maximum likelihood trees had similar topology, including H_V1 and C15orf27 families separating into discrete branches. A homology model of the VSD of H_V1 was constructed as described previously[17].

The C15orf27 clone was PCR-amplified from human cerebellum and subcloned into pcDNA3.1(+) expression vector (Invitrogen). The coding sequence of human H_V1 (*HVCN1*) was cloned into either pcDNA3.1(−) or pQBI25-fC3 (to make GFP–H_V1) vectors as described previously[16]. Site-directed mutants were created using the Stratagene QuikChange (Agilent) procedure according to the manufacturer's instructions. All the positive clones were sequenced to confirm the presence of the introduced mutation. HEK-293 or, more often COS-7 cells were grown to ~80% confluency in 35-mm cultures dishes, usually by seeding cells 1 day ahead of transfection. Cells were transfected with 0.4–0.5 μg of the appropriate cDNA using Lipofectamine 2000 (Invitrogen). After 6 h at 37 °C in 5% CO_2, the cells were trypsinized and replated onto glass cover slips at low density for patch clamp recording the following day. We selected green cells under fluorescence for recording. Patch-clamp methods were described previously[18].

The main pipette solution (also used externally) contained (in mM) 130 $TMACH_3SO_3$, 2 $MgCl_2$, 2 EGTA, 80 MES, titrated to pH 5.5 with ~20 TMAOH. In the pH 5.5 TMACl solution, TMACl replaced $TMACH_3SO_3$. Bath solutions at pH 7.0 had (mM) 90 $TMACH_3SO_3$ or TMACl, 3 $CaCl_2$, 1 EGTA, 100 BES, and 36–40 TMAOH. For experiments with Zn^{2+}, solutions contained PIPES buffer[37] without EGTA. Experiments were done at 20–25 °C. Currents are shown without leak correction. V_{rev} data were corrected for liquid junctions potentials measured in each solution[19].

31. Pei, J., Kim, B. H. & Grishin, N. V. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36,** 2295–2300 (2008).
32. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52,** 696–704 (2003).
33. Felsenstein, J. PROTPARS – Protein Sequence Parsimony Method http://cmgm.stanford.edu/phylip/protpars.html (1993).
34. Néron, B. *et al.* Mobyle: a new full web bioinformatics framework. *Bioinformatics* **25,** 3005–3011 (2009).
35. Chevenet, F., Brun, C., Bañuls, A. L., Jacq, B. & Christen, R. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* **7,** 439 (2006).
36. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23,** 127–128 (2007).
37. Cherny, V. V. & DeCoursey, T. E. pH-dependent inhibition of voltage-gated H^+ currents in rat alveolar epithelial cells by Zn^{2+} and other divalent cations. *J. Gen. Physiol.* **114,** 819–838 (1999).

©2011 Macmillan Publishers Limited. All rights reserved

CORRECTIONS & AMENDMENTS

CORRIGENDUM

doi:10.1038/nature10675

A draft genome of *Yersinia pestis* from victims of the Black Death

Kirsten I. Bos, Verena J. Schuenemann, G. Brian Golding, Hernán A. Burbano, Nicholas Waglechner, Brian K. Coombes, Joseph B. McPhee, Sharon N.DeWitte, Matthias Meyer, Sarah Schmedes, JamesWood, David J. D. Earn, D. Ann Herring, Peter Bauer, Hendrik N. Poinar & Johannes Krause

Nature **478,** 506–510 (2011)

In this Letter, the GenBank accession number was wrongly printed as SRA045745.1. It should be SRA045745. Supplementary Tables 3a and 4 have also been corrected.

©2011 Macmillan Publishers Limited. All rights reserved

CAREERS

COLUMN Expertise is overrated — sometimes naivety is best. **p.281**

FUNDING Website helps researchers to attract private support **p.281**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**



S. LUBENOW/JAI/CORBIS

The German government and private sector are pouring money into energy research and infrastructure, such as this solar-energy plant in North Frisia.

GERMANY

Renewables revolution

Europe's largest economy is boosting research on alternative energy sources and generating job opportunities.

BY QUIRIN SCHIERMEIER AND KATRIN KOHNERT

Marcus Bär is a prime target for headhunters. An electrical engineer, he leads an independent young-investigator group at the Helmholtz Centre Berlin for Materials and Energy (HZB), working on next-generation, high-performance thin-film solar cells. But his research experience has also attracted the interest of private-sector solar-energy companies throughout Germany, which are aggressively recruiting skilled scientists and engineers. Bär has received some tempting job offers, he says — but has opted to stay in his tenure-track academic position.

Solar research at the HZB is one of many alternative-energy initiatives to get under way in Germany in recent years. Energy research has long been a priority for the German government, but the meltdown in March at the Fukushima Daiichi nuclear plant in Japan gave it extra urgency. Before the disaster, nuclear energy had supplied some 23% of Germany's electricity demand. But immediately after the meltdown, the German government withdrew the licences of seven older nuclear reactors (see *Nature* **472,** 505; 2011), and later decided it would shut down the nine remaining reactors by the end of 2022. This has added impetus to plans for investment in energy research over the next several years.

In August, just weeks after announcing the nuclear phase-out, the government approved a €3.5-billion (US$4.7-billion) energy-research programme for 2011–14 — a 75% increase over the 2006–09 funding period. Of that, some €2.5 billion is earmarked for research on renewable energy and energy efficiency.

This means plenty of new jobs, for researchers at private companies as well as those at institutes that rely on government funding. Analysts with the German Institute of Economic Research in Berlin estimate that over the next decade, several hundred thousand skilled jobs could be generated in the renewable-energy sector, many for scientists and engineers. Companies and institutes will be recruiting from home and abroad; indeed, bringing in more foreign talent is a government policy.

VARIETY AND SCOPE

The job market spans a multitude of topics in basic and applied research, says Bernd Emonts, deputy head of the Institute of Energy and Climate Research at the Jülich Research Centre (FZJ), who oversees more than 100 people ▶

© 2011 Macmillan Publishers Limited. All rights reserved

working on fuel cells. Future work will focus not only on how to gain energy from renewable sources, but also on how to store and distribute electricity. The government's energy investment will allow the FZJ to expand nascent research on the potential of batteries and hydrogen for efficient energy storage. As energy research becomes a national priority, the centre will seek scientists and engineers from disciplines including material sciences, physical chemistry, chemical and material engineering and mathematics.



A. S. BUKVIĆ-SCHÄFER

"For me, the change to industry has been the right career move at the right time."
Aleksandra Saša Bukvić-Schäfer

"The market for new technologies, and the demand for young researchers who can develop it, will boom for many years to come," says Joachim Fuchs, programme manager for renewable energies at the Karlsruhe Institute of Technology (KIT), which will benefit from both the public funding and industry investment. In October, the institute opened the Battery and Electrochemistry Laboratory (BELLA) in partnership with chemical giant BASF, which is based in Ludwigshafen. Over the next 5 years, the KIT and BASF each plan to invest €6 million in energy-storage and electric-car research at the centre, which is about to advertise for 4–6 postdoctoral scientists.

Ideal candidates, says Jürgen Janek, a physical chemist and co-director of BELLA, would have a background in electrochemistry and skills in solid-state chemistry, nanotechnology or material sciences. But those with a general background in, for example, physics or inorganic chemistry could also find opportunities: employers can often train such candidates relatively quickly on the job, says Janek.

INDUSTRIAL OPPORTUNITIES

Renewable-energy technology in Germany was once the domain of small and medium-sized companies. But in September, one of the country's largest and most research-intensive companies — Siemens, headquartered in Munich — announced plans to withdraw from nuclear power and focus on the booming renewables sector. Michael Weinhold, chief technology officer of Siemens Energy in Erlangen, says that the company will focus on everything from developing small technical components to installing large facilities such as offshore wind farms and solar thermal power stations, and ultra-high-voltage technologies for optimized grids and improved energy transmission. Siemens Energy is constantly looking for well-trained young scientists and engineers, ideally with expertise in renewable energies, but also with more general bachelor's or master's degrees in electrical, chemical or mechanical engineering. Weinhold declined to provide exact hiring numbers.

Janek says that with so many groups and companies currently recruiting, "the market for highly skilled specialists is pretty much swept clean". Rapid growth in the field can mean a high turnover of personnel. The Fraunhofer Institute for Wind Energy and Systems Technology in Kassel and Bremerhaven, for example, hired around 50 specialists this year, but it also lost 15 scientists to industry. Some academics have recognized the opportunities. Aleksandra Saša Bukvić-Schäfer last year moved from the Fraunhofer Institute to SMA Solar Technology in Niestetal, a leading provider of photovoltaic technologies. "For me, the change to industry has been the right career move at the right time," she says. "I'm still involved in research on electrochemical energy storage, but my work here is more focused, less project-oriented — and better paid." Indeed, in response to increasing demand, salaries for industry-employed engineers have risen by more than 4% this year. Starting salaries range from an average of €42,000 for engineers with bachelor degrees to €52,000 for PhD holders.

GAP IN THE JOB MARKET

German technology companies, however, are facing a severe shortage of skilled engineers. The Association for German Engineers (VDI) in Düsseldorf reported that September saw a record number of private-sector vacancies — almost 100,000, many of them in renewable energy and electro-mobility. Bosch Solar Energy in Erfurt — a division of Bosch, one of the largest technology companies in Germany — is injecting €530 million into a centre for photovoltaics in Arnstadt. The centre will have up to 1,000 skilled technical and engineering jobs by 2012. "Electrical and mechanical engineers are particularly sought after," says Lars Funk, head of the VDI's profession and society division. "That doesn't mean that companies are cutting back on requirements with regards to applicants' skills. But employment opportunities are considerably improving."



VDI

"Electrical and mechanical engineers are particularly sought after."
Lars Funk

Next year, to make foreign hires more attractive to companies, the government is to lower the minimum annual salary that companies must pay specialists from outside the European Union (EU) before they can get a permanent work permit in Germany from €66,000 to €40,000.

Until now, foreign recruitment has been less successful than the government hoped. Companies are often unsure about the value of foreign degrees and qualifications, especially with regard to applicants from outside the EU. And employers must deal with work permits and pay relatively high salaries for skilled non-EU researchers. Academic jobs, which usually offer fixed-term appointments, are generally a better prospect for foreign scientists. Residence permits are easier to get — scientists can apply for a researcher's visa — and there is no minimum salary required to obtain a work permit.

DEGREE OF DIFFERENCE

To help to meet the demand, German research and applied-science universities offer more than 100 postgraduate renewable-energy programmes — covering aspects from basic science to energy economics and environmental management. Some programmes focus on specific technologies; for example, the Technical University of Berlin runs a master's programme in engineering for solar technology. Others, such as the University of Oldenburg's master of science in renewable energy, provide basic knowledge across different sectors. Most programmes are taught in English and encourage foreign students to apply. After the course, job prospects are good: most graduates of the Oldenburg programme go on to PhD research or take up positions in industry; others move on to careers in climate diplomacy or financial consulting, for instance, says Konrad Blum, an energy researcher at Oldenburg.

Prospective students should look into the curricula of the different programmes to see whether they provide qualifications relevant to the student's desired career path, says Theo Buehler, a career expert at the Wissenschaftsladen Bonn, a private 'science shop' that provides advice on the job market and qualification requirements in energy and environment sectors. To capitalize on the looming personnel shortage, the science shop recommends that students acquire solid core knowledge in a particular field, such as chemistry, before going on to specialize in renewable energy through a postgraduate programme. Language skills and flexibility are prerequisites, and business skills are an increasingly sought-after extra, says Buehler.

Bär knows that his own students and postdocs at the HZB are not immune to the lure of industry. "The best ones are the first to leave," he says. "So you've got to make sure that there is always enough fresh talent around. But looking at the many Asian, Middle Eastern and Latin American faces in my lectures, I'm not at all concerned that the well might run dry." ■

Quirin Schiermeier *is* Nature*'s Germany correspondent.* **Katrin Kohnert** *is a former* Nature *intern, based in Munich.*

© 2011 Macmillan Publishers Limited. All rights reserved

COLUMN

Unleash your inner dummy

There is something to be said for letting go of the mantle of expert, argues **Peter Fiske**.

C. REISINGER/SHUTTERSTOCK

In science, intellect and intelligence are valued above all else. Scientists spend years in graduate programmes, studying, teaching and researching, to become experts in their fields. Having invested so many years in developing our expertise, we naturally assume that it is the main thing that we can offer the world. Some assume that where we lack expertise, we cannot add any value.

The structure and culture of research tends to reinforce this idea. In papers, scientists reference only the leading experts. It is next to impossible for someone who is not recognized to get funding — no matter how good their grant application. Fear of failure causes many of us to avoid environments in which we are not experts. There may be no such thing as a stupid question in primary school, but in graduate school one can be made to feel quite uneasy about asking a basic question.

Ironically, always playing the expert can be limiting, in terms of both contributions to science and career options. Sometimes, playing the dummy can be liberating and help to reveal opportunities that would otherwise have been overlooked. Dummies ask questions that experts assume were answered long ago. Dummies explore subject areas in which they lack knowledge. Dummies listen more and talk less.

Becoming a dummy frees you from dogma. Developing expertise can often mean ingesting unquestioned assumptions and accepted facts. Such received beliefs can lead to unchallenged group decision-making and prevent a community from recognizing a path-breaking discovery — especially when it comes from someone outside the discipline.

In fact, many exciting scientific discoveries have arisen from a non-expert asking a basic question — consider, for example, the connection between the extinction of the dinosaurs and a meteorite impact, which was the brainchild of an experimental physicist, not a palaeontologist. Not coincidentally, experts in a field often meet these basic questions with a fair degree of hostility.

Embracing your inner dummy is also a powerful tool for communicating science. Many scientists assume that public hostility toward science stems from ignorance of the facts, and that science education is the remedy. But perhaps the answer is also more human engagement by scientists; sometimes they need to step out of the role of expert and seek to understand the audience's cultural and ethical perspectives.



Scientists' preoccupation with being an expert as they consider career options can be enormously limiting. Newly minted PhD holders can have a hard time seeing themselves as anything but researchers. The idea that they are experts in their subjects is familiar and comfortable. Less familiar is the notion that the transferable skills they have developed — such as analysis, research and model testing — may also make them adaptable problem-solvers, able to contribute in a wide range of career environments. As a result, many PhD holders limit their ambitions to a job within their discipline.

In reality, doing original scientific research in a graduate programme can prepare you for a variety of environments and roles. The years spent developing and testing a hypothesis, coping with limited resources and learning to use an array of complicated tools for gathering and analysing scientific data are excellent preparation for being innovative, resourceful and effective wherever you work. By letting go of the need to be an expert, you can approach unfamiliar problems and recognize the underlying connections with what you have already done. ■

Peter Fiske *is chief executive of PAX Water Technologies in Richmond, California, and author of* Put Your Science to Work *(American Geophysical Union, 2001).*

UNITED STATES

Doctorate awards drop

The number of science doctorates awarded in the United States has declined for the first time in eight years, according to a report by the US National Science Foundation (NSF). The NSF's annual survey, out in late November, revealed that US institutions conferred 25,820 science doctorates in 2009, but only 25,589 in 2010, a decrease of 0.89%. Between 2002 and 2009, the number climbed by 32%. However, 2009–10 saw small drops in agricultural sciences, molecular biology, neuroscience, Earth sciences and chemistry. Report author Mark Fiegener, a statistician at the NSF in Arlington, Virginia, says that the fall might have been caused by the uncertain economy, which is prompting science graduate students to stay in their programmes for longer, rather than claiming their degrees. "They may be staying in school until the economy looks better and they can get a position," says Fiegener.

UNITED KINGDOM

Manchester hiring

Despite austerity measures across Britain, the University of Manchester is creating more than 30 science faculty posts as part of a £20-million (US$31-million) recruitment campaign. It hopes to boost its performance in the Research Excellence Framework (REF) used to assess UK universities. The positions include 23 professorships and 7 lectureships, and are part of a move to establish more than 100 posts across all fields. A spokesperson says that the scheme will put Manchester ahead of institutions that are cutting back, so that it will score well in the 2014 REF and receive more funding. Manchester has a £40-million budget surplus this year as a result of cost-saving measures.

FUNDING

Gift system for science

An innovative website lets the public donate to scientific projects. SciFlies (http://sciflies.org), publicly launched on 22 November, lists and promotes peer-reviewed projects, giving researchers 6 months to reach funding targets. Creator David Fries, a marine scientist at the University of South Florida in St Petersburg, says that the site lets scientists raise money and build a community of supporters. Donations are free from university contract and overhead rates, says Fries, so researchers get a large share.

© 2011 Macmillan Publishers Limited. All rights reserved

EATING WITH INTEGRITY

A question of taste.

BY DAVID BERREBY

So, how was the big dinner party?

A complete disaster! Geoff and I haven't spoken since …

Not since *Friday*?

That's right. You know, sometimes we joke about divorce when we're arguing, but this time … it doesn't feel safe to kid about that. It's just so silent and grim. I think he thinks I've ruined his career.

Look, I'm sure you did your best. It was a lot of pressure. All three of his new colleagues, plus the Big Boss. I'm sure you did everything right.

I thought I had. I checked everyone for allergies. Religious covenants. Lipid and corticosterone metabolism. Medications, immunizations …

Well, you can't be too careful.

And I social-vetted the menu: everyone checked off everything. And there was nothing to offend anyone. Tempeh-salmon, made by our own 'bot, right here. All ingredients grown sterile, no contact with organic matter or pollutants …

You don't have to tell me. I know how careful we have to be. How careful you are. No one's been sick from your place in years, and how many of us can say that?

Thanks. I appreciate that, especially after what's happened. God, I don't even want to tell. It's just so mortifying.

I think you'd better, though.

Sigh. Well, then, from the top. The table is set and the containers are all sealed and ready to be opened, the dining tents are over the chairs …

You used the good ones?

Of course! Pressurized, sterilized, totally safe, near-transparent. The ones we got for our tenth anniversary. So there I am, all ready and Geoff is wandering around resterilizing forks because everything's perfect and he has nothing to do and then bing! In walks our first guest, who is also the Big Boss, and he's carrying several things that look like food boxes and I think, uh-oh. Then I think, well, of course, those must be gifts, maybe a floral hologram or something, because it can't be what I'm afraid it is. It just *can't be*!

Well, how could you expect *that*?

You know how this thing is spreading. But I didn't update before the dinner, so I managed to miss the day when Mr Big Boss became one of *them*.

We had three people at work do the same thing last week. It's like some damn cult!

"I hope it's not a problem," he says, with that hick accent, "but I've decided to *Eat With Integrity*." I mean, I about died. And the worst of it is, he's the *first guest*. Everyone else is going to think we approve of this kind of thing.

He's also Geoff's boss. You didn't have much choice.

I know, but I was just so … appalled. And disgusted. And I had no idea what to do, because we don't have the equipment.

You mean it *wasn't cooked*?

Oh, no, of course it wasn't cooked. Mr Big isn't just a new-minted Integretist, he's a hard-core Integretist. The meat has to be fresh as can be. He harvests himself, he cooks himself, it has to be done within the hour.



JACEY

You know, I just realized, in a way it's … well, it's sort of like what our great-grandparents ate, isn't it? Kind of like … meat?

Exactly the problem! It's not tofurkey or FauxPorc, I can assure you. It's red and bloody and it smells of iron and something else …

I'm sorry, I think I'm going to be sick.

You asked me to tell you!

Oh, God. To think, in your own home!

Right, so there he is, and he forceps out this … lump of flesh, this oozing … *thing* like it's the Dark Ages and he looks at me cool as a cucumber and asks me if he can use the stove. And that's when most of the others walked in. I don't know what was worse — the disgust as they looked at us, or the pity. So, of course, into this conversational vacuum leaps our newly converted Integretist. How he can talk! How he'd been thinking about Eating With Integrity for years, how much sense it made, what with all the contaminants and the epidemics, you couldn't be too careful, blah blah blah. How it was all well and good to grow your own foods and use sterile hydroponics but you couldn't be too sure, and anyway, we evolved to eat meat, and the fact that there are 10 billion of us on Earth doesn't change our basic anatomy, yak yak yak. Stem-cell technology is perfected now, and you can grow whatever you like, they have scaffolds for anything, muscle, liver, sweetbreads …

Oh, my. No one spoke up?

He's the boss! No one dared. And I think they were all just stunned to see that this movement's spreading into their own lives, into their own office. It's not just for recovering plague victims and famine survivors any more.

You know, I saw something the other night about that. The Integretists are getting up to 20% in some cities. Once it costs nothing to clone yourself, a certain kind of person thinks, well, why not?

Meanwhile the stove is doing what it can, trying to cook this … this *lump*. So it starts to … smell, you know? And a couple of the more sensitive people, Geoff's brother, for one, they excuse themselves. And that sets off our guest.

It's the safest meat there is, he says. People didn't have to agree, but they shouldn't feel free to be downright rude.

Bit of a buzz-kill, eh?

… So then people force themselves to sit down, and they're tented up, and their containers unseal, but … God, I can still see it! No one eats a thing. We're all just too disgusted and perplexed, so the only sound at the table is my one happy guest, Eating With Integrity, happily chewing away on what he keeps insisting is the only 100% safe and healthy food left: his own flesh, cloned and grown in a vat next to his coffee machine. I can still see him there, having harvested himself and cooked himself, now eating himself; holding up his fork to our averted eyes, smiling and saying, now this, *this* is *Eating With Integrity*! ■

NATURE.COM
Follow Futures on Facebook at:
go.nature.com/mtoodm

A science writer, **David Berreby** *writes the Mind Matters blog at bigthink.com/blogs/Mind-Matters and is the author of* Us and Them: The Science of Identity. *He lives in Brooklyn, New York.*

© 2011 Macmillan Publishers Limited. All rights reserved

GSK-3α/β kinases and amyloid production *in vivo*

ARISING FROM C. J. Phiel, C. A. Wilson, V. M.-Y. Lee & P. S. Klein *Nature* **423,** 435–439 (2003)

A major unresolved issue in Alzheimer's disease is identifying the mechanisms that regulate proteolytic processing of amyloid precursor protein (APP)—glycogen synthase kinase-3 (GSK-3) isozymes are thought to be important in this regulation. Phiel *et al.*[1] proposed that GSK-3α, but not GSK-3β, controls production of amyloid[1]. We analysed the proteolytic processing of mouse and human APP in mouse brain *in vivo* in five different genetic and viral models. Our data do not yield evidence for either GSK-3α-mediated or GSK-3β-mediated control of APP processing in brain *in vivo*.

GSK-3β is believed to be central to the pathogenesis of Alzheimer's disease, linking amyloid and tau pathology[2–5]. Unlike GSK-3β, neither physiological functions nor pathological roles of GSK-3α are well explored[6]. To analyse the function of both GSK-3 kinases in brain, we generated mice that were completely deficient in GSK-3α (denoted as *Gsk3a*^{KO}) as well as mice with neuron-specific GSK-3α or GSK-3β deficiency using Cre/*loxP* as for presenilin 1 (ref. 7).

The contribution of GSK-3α to the processing of APP was analysed by Phiel *et al.*[1] by silencing with short interfering RNA in cells and by pharmaceutical inhibition with Li^+ ions (an inhibitor of both GSK-3 isozymes). We approached this problem genetically by generating two strains of GSK-3α-deficient mice, denoted as *Gsk3a*^{KO} and *Gsk3a*$^{n-/-}$ (complete and neuron-specific GSK-3α deficiency, respectively). In *Gsk3a*^{KO} mice, the gene was inactivated completely in all tissues, demonstrated by genotyping, immunohistochemistry and western blotting (data not shown). In the brains of *Gsk3a*$^{n-/-}$ mice, some residual activity was expected because GSK-3 enzymes are expressed in non-neuronal cell types in the CNS[5,8] (Fig. 1). GSK-3α deficiency had no evident impact on GSK-3β levels, demonstrating their independent regulation (Fig. 1).

First, we analysed biochemically the proteolytic processing of endogenous mouse APP in brains of adult *Gsk3a*$^{n-/-}$ mice by measuring the APP metabolites in total brain extracts by validated methods[5,7,9–11]. Neuronal deficiency of GSK-3α did not affect levels of immature and mature APP, APP phosphorylated at T668 (pT668), secreted ectodomain APP, nor of carboxy-terminal fragments (Fig. 1a). Levels of mouse amyloid peptides in the brain, measured by specific enzyme-linked immunosorbent assay (ELISA), were unaffected in *Gsk3a*$^{n-/-}$ mice relative to age- and gender-matched control mice with floxed *Gsk3a* genes but not expressing Cre recombinase (Fig. 1a).

We went on to administer AAV-APP.SLA viral vectors (adeno-associated viral vector expressing triple mutant APP with Swedish, London and Austrian mutations) by intra-hippocampal injection into *Gsk3a*$^{n-/-}$ mice, to analyse the processing of human mutant APP.SLA (ref. 9). At 3 weeks after injection, biochemical analysis of hippocampal extracts of AAV-APP.SLA-injected *Gsk3a*$^{n-/-}$ mice revealed no significant differences for any of the APP metabolites (Fig. 1b).

We continued to analyse mice with a complete deficiency in GSK-3α generated serendipitously during the expansion of the *Gsk3a*$^{n-/-}$

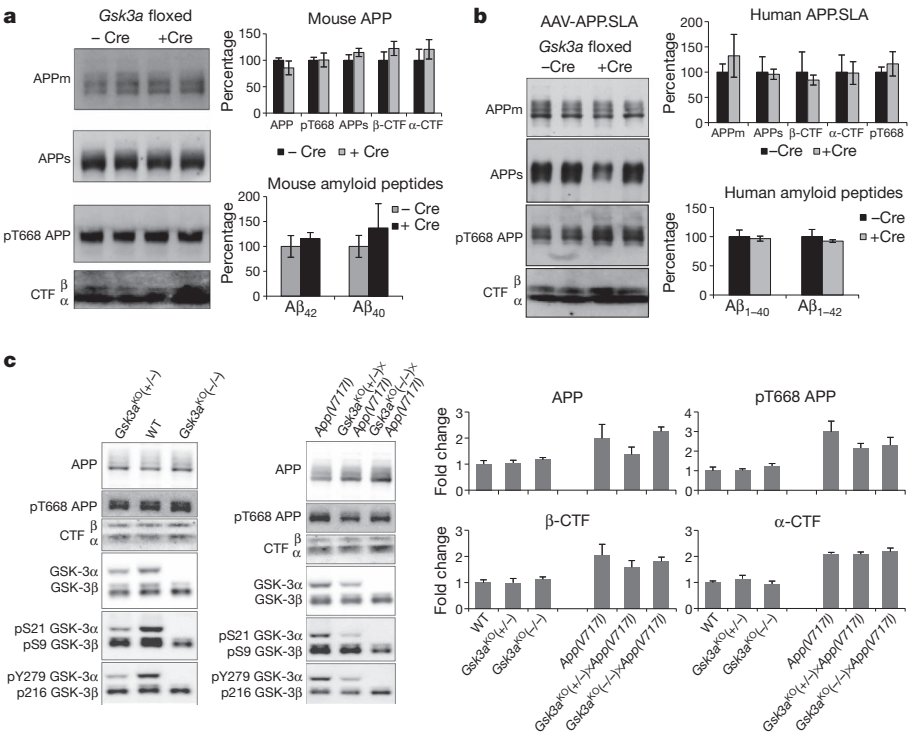


Figure 1 | APP processing in the brains of *Gsk3a*$^{n-/-}$ and *Gsk3a*^{KO} mice. **a**, Biochemical analysis of murine APP metabolites in total brain homogenates by western blotting. Mouse amyloid peptides were measured by specific ELISA. APPm, membrane-bound APP; APPs, secreted APP; Cre, Cre recombinase; CTF, C-terminal fragments; floxed, recombinant genes with inserted *loxP* elements[7]. **b**, Biochemical analysis of hippocampi at 3 weeks after intracerebral injection of AAV-APP.SLA. Human amyloid peptides were measured by specific ELISA[7]. **c**, Analysis of murine APP metabolites in the brains of *Gsk3a*^{KO} mice and of human APP metabolites in *Gsk3a*^{KO}×*App(V717I)* bigenic mice, heterozygous (+/−) or homozygous (−/−) for GSK-3α deficiency. All data are mean ± s.e.m.

©2011 Macmillan Publishers Limited. All rights reserved

colony (data not shown). In addition, we generated bigenic mice by crossing the *Gsk3a*^{KO} mice with our *App(V717I)* mice, a validated model for amyloid pathology[5,7,9–11]. We carried out a biochemical analysis of metabolites of mouse APP and human mutant APP in brains of *Gsk3a*^{KO} mice and *Gsk3a*^{KO}×*App(V717I)* bigenic mice, respectively. None of the APP metabolites derived from either mouse or human APP was affected by the complete deficiency of GSK-3α (Fig. 1c).

We subsequently extended the study to the GSK-3β isozyme by generating *Gsk3b*$^{n-/-}$ mice with a neuron-specific deficiency of GSK-3β. GSK-3β deficiency had no evident impact on the expression of the GSK-3α isozyme, further corroborating their independent regulation (Fig. 2). Notably, no significant deviation was noted in the brain levels of mouse APP metabolites in *Gsk3b*$^{n-/-}$ mice (Fig. 2).

In addition, we attempted to generate mice deficient in both GSK-3 isozymes in their central neurons. The combination of mice containing both floxed *Gsk3* genes and the Thy1-Cre recombinase transgene yielded 354 offspring over a time span of 16 months that were all genotyped for the five genes of interest: wild-type and floxed GSK-3 isozymes and Thy1-Cre. None of the pups contained the wanted combination with both *Gsk3* genes recombined homozygously. The outcome proved, not unexpectedly, that complete deficiency of GSK-3 activity in central neurons is lethal for mice.

We conclude that the GSK-3 isozymes do not contribute significantly to the processing of APP in mouse brain *in vivo*. Of interest, the combination of *Gsk3b*$^{n-/-}$ with *App(V717I)* did not yield viable offspring (data not shown), in contrast to viable offspring yielded by the *Gsk3a*^{KO}×*App(V717I)* combination (Fig. 1c). This underlines the substantial functional difference of the GSK-3 isozymes in brain; a difference that we believe is not, however, related to APP processing. The combined outcome substantiates the notion that GSK-3β is physiologically and pathologically the dominant isozyme, notwithstanding the fact that both are activated by amyloid[5]. The latter point testifies that both kinases contribute, downstream of APP, to tauopathy and cognitive defects in bigenic and viral mouse models[3,5,12,13] (data not shown) and by extrapolation in human disease[2,4,13].

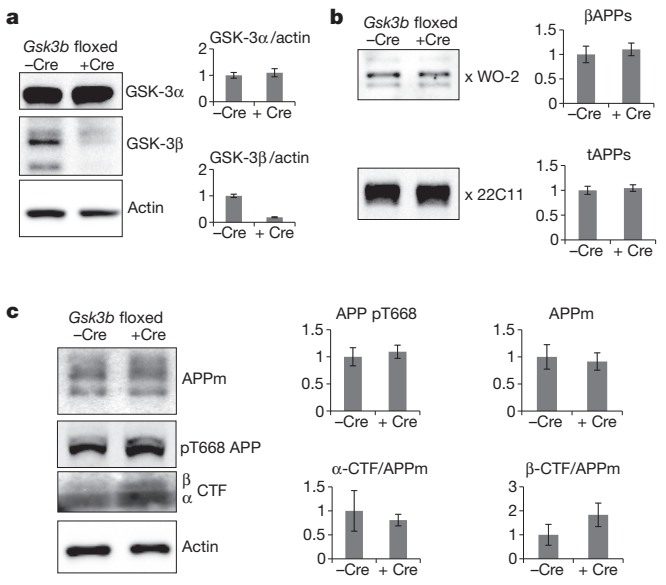


Figure 2 | APP processing in the brains of *Gsk3b*$^{n-/-}$ mice. **a**, Western blot of GSK-3 in brain homogenates of mice with floxed *Gsk3b* genes, with and without expression of Cre recombinase. **b**, **c**, Biochemical analysis of proteolytic processing of mouse APP in the brains of *Gsk3b*$^{n-/-}$ mice. All data are mean ± s.e.m. The *y* axes of graphs in **a**–**c** show relative percentage. βAPPs, β-secretase cleaved secreted APP; tAPPs, total secreted APP; WO-2, monoclonal antibody specific for the amyloid sequence in APP and amyloid peptides; 22C11, monoclonal antibodies specific for N-terminus of APP.

Tomasz Jaworski[1], Ilse Dewachter[1]†, Benoit Lechat[1], Maarten Gees[1]†, Anna Kremer[1], David Demedts[1], Peter Borghgraef[1], Herman Devijver[1], Seb Kügler[2], Satish Patel[3], Jim R. Woodgett[3] & Fred Van Leuven[1]

[1]Experimental Genetics Group, KU Leuven, Leuven, Belgium.
e-mail: fred.vanleuven@med.kuleuven.be
[2]Department of Neurology, Georg-August-University, Göttingen, Germany.
[3]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada.
†Present addresses: FARL-Institute of Neurosciences, UC Louvain, B-1000 Brussels, Belgium (I.D.); Laboratory of Ion Channel Research, KU Leuven, B-3000 Leuven, Belgium (M.G.).

Received 1 July; accepted 15 September 2011.

1. Phiel, C. J., Wilson, C. A., Lee, V. M. & Klein, P. S. GSK-3α regulates production of Alzheimer's disease amyloid-β peptides. *Nature* **423,** 435–439 (2003).
2. Ishiguro, K. *et al.* Glycogen synthase kinase 3β is identical to tau protein kinase I generating several epitopes of paired helical filaments. *FEBS Lett.* **325,** 167–172 (1993).
3. Spittaels, K. *et al.* GSK3β phosphorylates protein tau and rescues the axonopathy in the central nervous system of human four-repeat tau transgenic mice. *J. Biol. Chem.* **275,** 41340–41349 (2000).
4. Takashima, A. GSK-3 is essential in the pathogenesis of Alzheimer's disease. *J. Alzheimers Dis.* **9** (Suppl.), 309–317 (2006).
5. Terwel, D. *et al.* Amyloid activates GSK-3β to aggravate neuronal tauopathy in bigenic mice. *Am. J. Pathol.* **172,** 786–798 (2008).
6. Kaidanovich-Beilin, O. *et al.* Abnormalities in brain structure and behavior in GSK-3α mutant mice. *Mol Brain.* **2,** 35 (2009).
7. Dewachter, I. *et al.* Neuronal deficiency of presenilin 1 inhibits amyloid plaque formation and corrects hippocampal long-term potentiation but not a cognitive defect of amyloid precursor protein [V717I] transgenic mice. *J. Neurosci.* **22,** 3445–3453 (2002).
8. Perez-Costas, E., Gandy, J. C., Melendez-Ferro, M., Roberts, R. C. & Bijur, G. N. Light and electron microscopy study of glycogen synthase kinase-3β in the mouse brain. *PLoS ONE* **5,** e8911 (2010).
9. Jaworski, T. *et al.* AAV-tau mediates pyramidal neurodegeneration by cell-cycle re-entry without neurofibrillary tangle formation in wild-type mice. *PLoS ONE* **4,** e7280 (2009).
10. De Strooper, B. *et al.* Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein. *Nature* **391,** 387–390 (1998).
11. Moechars, D. *et al.* Early phenotypic changes in transgenic mice that overexpress different mutants of amyloid precursor protein in brain. *J. Biol. Chem.* **274,** 6483–6492 (1999).
12. Tanghe, A. *et al.* Pathological hallmarks, clinical parallels, and value for drug testing in Alzheimer's disease of the APP[V717I] London transgenic mouse model. *Int. J. Alzheimers Dis.* **2010,** 417314 (2010).
13. Jaworski, T. *et al.* Alzheimer's disease: old problem, new views from transgenic and viral models. *Biochim. Biophys. Acta* **1802,** 808–818 (2010).

Author Contributions Generation of transgenic and knockout mice, B.L., D.D., H.D., S.P. and J.R.W.; generation of AAV vectors and intracerebral injections, T.J., D.D. and S.K.; brain analysis, T.J., I.D., B.L., M.G., A.K. and P.B.; design of experiments, data analysis and figures, T.J., I.D., P.B. and F.V.L.; writing of manuscript, T.J. and F.V.L. T.J., I.D. and B.L. contributed equally to this work.

Competing financial interests: declared none.

doi:10.1038/nature10615

©2011 Macmillan Publishers Limited. All rights reserved

Phiel *et al.* reply

REPLYING TO T. Jaworski *et al. Nature* **480,** doi:10.1038/nature10615 (2011)

GSK-3 has been implicated in the pathogenesis of Alzheimer's disease through regulation of tau phosphorylation, cellular responses to amyloid-β and processing of amyloid precursor protein (APP) to amyloid-β. We previously reported[1] that inhibition of GSK-3 reduces the accumulation of $A\beta_{40}$ and $A\beta_{42}$ (amyloid-ß peptides of 40 and 42 amino acids) in mouse brain and cell culture and that knockdown of *Gsk3a* reduces amyloid-β accumulation in CHO cells, indicating that *Gsk3a* contributes to the processing of APP in this context[1]. At about the same time, others reported that knockdown of *Gsk3b* reduces amyloid-β production and overexpression of active *Gsk3b* enhances amyloid-β production[2–4]. A reasonable explanation for these findings, as suggested previously[2,5,6], was that both of these highly similar enzymes contribute to APP processing and their respective contributions depend on cell type, relative abundance and assay conditions. Jaworski *et al.*[7] now show that APP can be processed in mouse brain in the absence of either *Gsk3a* or neuronal *Gsk3b*.

However, *Gsk3a* and *Gsk3b* are highly homologous and frequently redundant genes[8,9]. For this reason, it is not possible to exclude a role for GSK-3 with single gene knockouts. Furthermore, structurally and mechanistically diverse GSK-3 inhibitors reduce amyloid-β accumulation *in vivo* and in cell culture, including lithium, kenpaullone, bisindolylmaleimide I, FRAT peptide, the TDZD-related NP12 and kinase-dead GSK-3 (refs 1–6, 10, 11). These compounds inhibit both GSK-3α and GSK-3β and therefore get around the issue of redundancy. Although off-target effects for these diverse GSK-3 inhibitors could, formally, explain their effects on APP processing, a far more plausible explanation is that these compounds act through GSK-3 inhibition.

Jaworski *et al.*[7] show that APP can be processed in mouse brain in the absence of either *Gsk3a* or neuronal *Gsk3b*, but do not examine redundant functions of *Gsk3a* and *Gsk3b* in APP processing and do not consider the substantial body of work showing that acute inhibition of GSK-3 reduces amyloid-β accumulation *in vivo*. The simplest explanation for these findings[2,5,6] is that both *Gsk3a* and *Gsk3b* can contribute to APP processing and that inhibition of GSK-3 reduces amyloid-β in Alzheimer's disease models.

Christopher J. Phiel[1], Christina A. Wilson[2], Virginia M.-Y. Lee[3] & Peter S. Klein[4]

[1]Center for Molecular and Human Genetics, The Research Institute at Nationwide Children's Hospital, Columbus, Ohio 43205, USA.
[2]Department of Neurology, 3 West Gates Building, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA.
[3]Center for Neurodegenerative Disease Research, Department of Pathology and Laboratory Medicine, 3rd Floor Maloney Building, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA.
[4]Department of Medicine, Division of Hematology–Oncology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104-6148, USA.
e-mail: pklein@mail.med.upenn.edu

1. Phiel, C. J., Wilson, C. A., Lee, V. M. & Klein, P. S. GSK-3α regulates production of Alzheimer's disease amyloid-β peptides. *Nature* **423,** 435–439 (2003).
2. Ryder, J. *et al.* Divergent roles of GSK3 and CDK5 in APP processing. *Biochem. Biophys. Res. Commun.* **312,** 922–929 (2003).
3. Li, B. *et al.* FRAT1 peptide decreases Aβ production in $swAPP_{751}$ cells. *FEBS Lett.* **553,** 347–350 (2003).
4. Su, Y. *et al.* Lithium, a common drug for bipolar disorder treatment, regulates amyloid-β precursor protein processing. *Biochemistry* **43,** 6899–6908 (2004).
5. Huang, H. C. & Klein, P. S. Multiple roles for glycogen synthase kinase-3 as a drug target in Alzheimer's disease. *Curr. Drug Targets* **7,** 1389–1397 (2006).
6. Huang, H. C., O'Brien, W. T. & Klein, P. S. Targeting glycogen synthase kinase-3 in Alzheimer's disease. *Drug Discov. Today Ther. Strateg.* **3,** 613–619 (2006).
7. Jaworski, T. *et al.* GSK-3α/β kinases and amyloid production *in vivo*. *Nature* **480,** doi:10.1038/nature10615 (2011).
8. Doble, B. W. & Woodgett, J. R. GSK-3: tricks of the trade for a multi-tasking kinase. *J. Cell Sci.* **116,** 1175–1186 (2003).
9. Doble, B. W., Patel, S., Wood, G. A., Kockeritz, L. K. & Woodgett, J. R. Functional redundancy of GSK-3α and GSK-3β in Wnt/β-catenin signaling shown by using an allelic series of embryonic stem cell lines. *Dev. Cell* **12,** 957–971 (2007).
10. Sun, X. *et al.* Lithium inhibits amyloid secretion in COS7 cells transfected with amyloid precursor protein C100. *Neurosci. Lett.* **321,** 61–64 (2002).
11. Serenó, L. *et al.* A novel GSK-3β inhibitor reduces Alzheimer's pathology and rescues neuronal loss *in vivo*. *Neurobiol. Dis.* **35,** 359–367 (2009).

doi:10.1038/nature10616

©2011 Macmillan Publishers Limited. All rights reserved

GSK-3α/β kinases and amyloid production *in vivo*

ARISING FROM C. J. Phiel, C. A. Wilson, V. M.-Y. Lee & P. S. Klein *Nature* **423,** 435–439 (2003)

A major unresolved issue in Alzheimer's disease is identifying the mechanisms that regulate proteolytic processing of amyloid precursor protein (APP)—glycogen synthase kinase-3 (GSK-3) isozymes are thought to be important in this regulation. Phiel *et al.*[1] proposed that GSK-3α, but not GSK-3β, controls production of amyloid[1]. We analysed the proteolytic processing of mouse and human APP in mouse brain *in vivo* in five different genetic and viral models. Our data do not yield evidence for either GSK-3α-mediated or GSK-3β-mediated control of APP processing in brain *in vivo*.

GSK-3β is believed to be central to the pathogenesis of Alzheimer's disease, linking amyloid and tau pathology[2–5]. Unlike GSK-3β, neither physiological functions nor pathological roles of GSK-3α are well explored[6]. To analyse the function of both GSK-3 kinases in brain, we generated mice that were completely deficient in GSK-3α (denoted as *Gsk3a*^{KO}) as well as mice with neuron-specific GSK-3α or GSK-3β deficiency using Cre/*loxP* as for presenilin 1 (ref. 7).

The contribution of GSK-3α to the processing of APP was analysed by Phiel *et al.*[1] by silencing with short interfering RNA in cells and by pharmaceutical inhibition with Li^+ ions (an inhibitor of both GSK-3 isozymes). We approached this problem genetically by generating two strains of GSK-3α-deficient mice, denoted as *Gsk3a*^{KO} and *Gsk3a*$^{n-/-}$ (complete and neuron-specific GSK-3α deficiency, respectively). In *Gsk3a*^{KO} mice, the gene was inactivated completely in all tissues, demonstrated by genotyping, immunohistochemistry and western blotting (data not shown). In the brains of *Gsk3a*$^{n-/-}$ mice, some residual activity was expected because GSK-3 enzymes are expressed in non-neuronal cell types in the CNS[5,8] (Fig. 1). GSK-3α deficiency had no evident impact on GSK-3β levels, demonstrating their independent regulation (Fig. 1).

First, we analysed biochemically the proteolytic processing of endogenous mouse APP in brains of adult *Gsk3a*$^{n-/-}$ mice by measuring the APP metabolites in total brain extracts by validated methods[5,7,9–11]. Neuronal deficiency of GSK-3α did not affect levels of immature and mature APP, APP phosphorylated at T668 (pT668), secreted ectodomain APP, nor of carboxy-terminal fragments (Fig. 1a). Levels of mouse amyloid peptides in the brain, measured by specific enzyme-linked immunosorbent assay (ELISA), were unaffected in *Gsk3a*$^{n-/-}$ mice relative to age- and gender-matched control mice with floxed *Gsk3a* genes but not expressing Cre recombinase (Fig. 1a).

We went on to administer AAV-APP.SLA viral vectors (adeno-associated viral vector expressing triple mutant APP with Swedish, London and Austrian mutations) by intra-hippocampal injection into *Gsk3a*$^{n-/-}$ mice, to analyse the processing of human mutant APP.SLA (ref. 9). At 3 weeks after injection, biochemical analysis of hippocampal extracts of AAV-APP.SLA-injected *Gsk3a*$^{n-/-}$ mice revealed no significant differences for any of the APP metabolites (Fig. 1b).

We continued to analyse mice with a complete deficiency in GSK-3α generated serendipitously during the expansion of the *Gsk3a*$^{n-/-}$

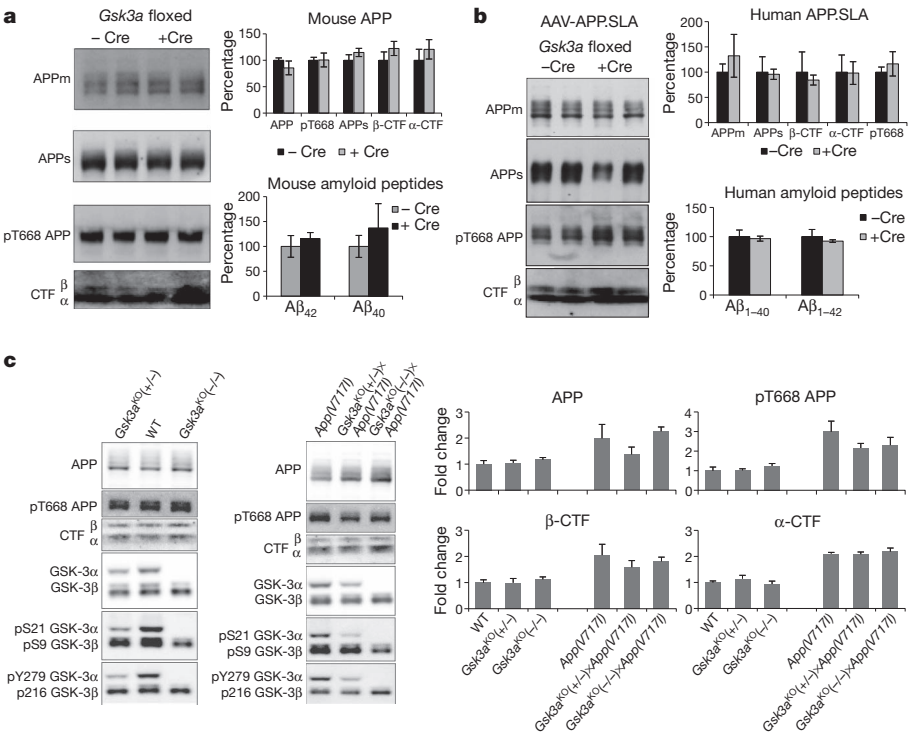


Figure 1 | APP processing in the brains of *Gsk3a*$^{n-/-}$ and *Gsk3a*^{KO} mice. **a**, Biochemical analysis of murine APP metabolites in total brain homogenates by western blotting. Mouse amyloid peptides were measured by specific ELISA. APPm, membrane-bound APP; APPs, secreted APP; Cre, Cre recombinase; CTF, C-terminal fragments; floxed, recombinant genes with inserted *loxP* elements[7]. **b**, Biochemical analysis of hippocampi at 3 weeks after intracerebral injection of AAV-APP.SLA. Human amyloid peptides were measured by specific ELISA[7]. **c**, Analysis of murine APP metabolites in the brains of *Gsk3a*^{KO} mice and of human APP metabolites in *Gsk3a*^{KO}×*App(V717I)* bigenic mice, heterozygous (+/−) or homozygous (−/−) for GSK-3α deficiency. All data are mean ± s.e.m.

©2011 Macmillan Publishers Limited. All rights reserved

colony (data not shown). In addition, we generated bigenic mice by crossing the *Gsk3a*^{KO} mice with our *App(V717I)* mice, a validated model for amyloid pathology[5,7,9–11]. We carried out a biochemical analysis of metabolites of mouse APP and human mutant APP in brains of *Gsk3a*^{KO} mice and *Gsk3a*^{KO}×*App(V717I)* bigenic mice, respectively. None of the APP metabolites derived from either mouse or human APP was affected by the complete deficiency of GSK-3α (Fig. 1c).

We subsequently extended the study to the GSK-3β isozyme by generating *Gsk3b*$^{n-/-}$ mice with a neuron-specific deficiency of GSK-3β. GSK-3β deficiency had no evident impact on the expression of the GSK-3α isozyme, further corroborating their independent regulation (Fig. 2). Notably, no significant deviation was noted in the brain levels of mouse APP metabolites in *Gsk3b*$^{n-/-}$ mice (Fig. 2).

In addition, we attempted to generate mice deficient in both GSK-3 isozymes in their central neurons. The combination of mice containing both floxed *Gsk3* genes and the Thy1-Cre recombinase transgene yielded 354 offspring over a time span of 16 months that were all genotyped for the five genes of interest: wild-type and floxed GSK-3 isozymes and Thy1-Cre. None of the pups contained the wanted combination with both *Gsk3* genes recombined homozygously. The outcome proved, not unexpectedly, that complete deficiency of GSK-3 activity in central neurons is lethal for mice.

We conclude that the GSK-3 isozymes do not contribute significantly to the processing of APP in mouse brain *in vivo*. Of interest, the combination of *Gsk3b*$^{n-/-}$ with *App(V717I)* did not yield viable offspring (data not shown), in contrast to viable offspring yielded by the *Gsk3a*^{KO}×*App(V717I)* combination (Fig. 1c). This underlines the substantial functional difference of the GSK-3 isozymes in brain; a difference that we believe is not, however, related to APP processing. The combined outcome substantiates the notion that GSK-3β is physiologically and pathologically the dominant isozyme, notwithstanding the fact that both are activated by amyloid[5]. The latter point testifies that both kinases contribute, downstream of APP, to tauopathy and cognitive defects in bigenic and viral mouse models[3,5,12,13] (data not shown) and by extrapolation in human disease[2,4,13].

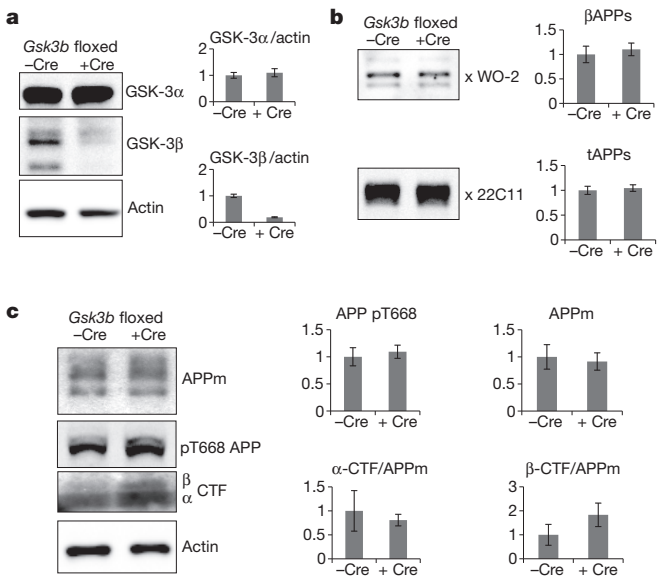


Figure 2 | APP processing in the brains of *Gsk3b*$^{n-/-}$ mice. **a**, Western blot of GSK-3 in brain homogenates of mice with floxed *Gsk3b* genes, with and without expression of Cre recombinase. **b**, **c**, Biochemical analysis of proteolytic processing of mouse APP in the brains of *Gsk3b*$^{n-/-}$ mice. All data are mean ± s.e.m. The *y* axes of graphs in **a**–**c** show relative percentage. βAPPs, β-secretase cleaved secreted APP; tAPPs, total secreted APP; WO-2, monoclonal antibody specific for the amyloid sequence in APP and amyloid peptides; 22C11, monoclonal antibodies specific for N-terminus of APP.

Tomasz Jaworski[1], Ilse Dewachter[1]†, Benoit Lechat[1], Maarten Gees[1]†, Anna Kremer[1], David Demedts[1], Peter Borghgraef[1], Herman Devijver[1], Seb Kügler[2], Satish Patel[3], Jim R. Woodgett[3] & Fred Van Leuven[1]

[1]Experimental Genetics Group, KU Leuven, Leuven, Belgium.
e-mail: fred.vanleuven@med.kuleuven.be
[2]Department of Neurology, Georg-August-University, Göttingen, Germany.
[3]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada.
†Present addresses: FARL-Institute of Neurosciences, UC Louvain, B-1000 Brussels, Belgium (I.D.); Laboratory of Ion Channel Research, KU Leuven, B-3000 Leuven, Belgium (M.G.).

Received 1 July; accepted 15 September 2011.

1. Phiel, C. J., Wilson, C. A., Lee, V. M. & Klein, P. S. GSK-3α regulates production of Alzheimer's disease amyloid-β peptides. *Nature* **423,** 435–439 (2003).
2. Ishiguro, K. *et al.* Glycogen synthase kinase 3β is identical to tau protein kinase I generating several epitopes of paired helical filaments. *FEBS Lett.* **325,** 167–172 (1993).
3. Spittaels, K. *et al.* GSK3β phosphorylates protein tau and rescues the axonopathy in the central nervous system of human four-repeat tau transgenic mice. *J. Biol. Chem.* **275,** 41340–41349 (2000).
4. Takashima, A. GSK-3 is essential in the pathogenesis of Alzheimer's disease. *J. Alzheimers Dis.* **9** (Suppl.), 309–317 (2006).
5. Terwel, D. *et al.* Amyloid activates GSK-3β to aggravate neuronal tauopathy in bigenic mice. *Am. J. Pathol.* **172,** 786–798 (2008).
6. Kaidanovich-Beilin, O. *et al.* Abnormalities in brain structure and behavior in GSK-3α mutant mice. *Mol Brain.* **2,** 35 (2009).
7. Dewachter, I. *et al.* Neuronal deficiency of presenilin 1 inhibits amyloid plaque formation and corrects hippocampal long-term potentiation but not a cognitive defect of amyloid precursor protein [V717I] transgenic mice. *J. Neurosci.* **22,** 3445–3453 (2002).
8. Perez-Costas, E., Gandy, J. C., Melendez-Ferro, M., Roberts, R. C. & Bijur, G. N. Light and electron microscopy study of glycogen synthase kinase-3β in the mouse brain. *PLoS ONE* **5,** e8911 (2010).
9. Jaworski, T. *et al.* AAV-tau mediates pyramidal neurodegeneration by cell-cycle re-entry without neurofibrillary tangle formation in wild-type mice. *PLoS ONE* **4,** e7280 (2009).
10. De Strooper, B. *et al.* Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein. *Nature* **391,** 387–390 (1998).
11. Moechars, D. *et al.* Early phenotypic changes in transgenic mice that overexpress different mutants of amyloid precursor protein in brain. *J. Biol. Chem.* **274,** 6483–6492 (1999).
12. Tanghe, A. *et al.* Pathological hallmarks, clinical parallels, and value for drug testing in Alzheimer's disease of the APP[V717I] London transgenic mouse model. *Int. J. Alzheimers Dis.* **2010,** 417314 (2010).
13. Jaworski, T. *et al.* Alzheimer's disease: old problem, new views from transgenic and viral models. *Biochim. Biophys. Acta* **1802,** 808–818 (2010).

Author Contributions Generation of transgenic and knockout mice, B.L., D.D., H.D., S.P. and J.R.W.; generation of AAV vectors and intracerebral injections, T.J., D.D. and S.K.; brain analysis, T.J., I.D., B.L., M.G., A.K. and P.B.; design of experiments, data analysis and figures, T.J., I.D., P.B. and F.V.L.; writing of manuscript, T.J. and F.V.L. T.J., I.D. and B.L. contributed equally to this work.

Competing financial interests: declared none.

doi:10.1038/nature10615

©2011 Macmillan Publishers Limited. All rights reserved

Phiel *et al.* reply

REPLYING TO T. Jaworski *et al. Nature* **480,** doi:10.1038/nature10615 (2011)

GSK-3 has been implicated in the pathogenesis of Alzheimer's disease through regulation of tau phosphorylation, cellular responses to amyloid-β and processing of amyloid precursor protein (APP) to amyloid-β. We previously reported[1] that inhibition of GSK-3 reduces the accumulation of $A\beta_{40}$ and $A\beta_{42}$ (amyloid-ß peptides of 40 and 42 amino acids) in mouse brain and cell culture and that knockdown of *Gsk3a* reduces amyloid-β accumulation in CHO cells, indicating that *Gsk3a* contributes to the processing of APP in this context[1]. At about the same time, others reported that knockdown of *Gsk3b* reduces amyloid-β production and overexpression of active *Gsk3b* enhances amyloid-β production[2–4]. A reasonable explanation for these findings, as suggested previously[2,5,6], was that both of these highly similar enzymes contribute to APP processing and their respective contributions depend on cell type, relative abundance and assay conditions. Jaworski *et al.*[7] now show that APP can be processed in mouse brain in the absence of either *Gsk3a* or neuronal *Gsk3b*.

However, *Gsk3a* and *Gsk3b* are highly homologous and frequently redundant genes[8,9]. For this reason, it is not possible to exclude a role for GSK-3 with single gene knockouts. Furthermore, structurally and mechanistically diverse GSK-3 inhibitors reduce amyloid-β accumulation *in vivo* and in cell culture, including lithium, kenpaullone, bisindolylmaleimide I, FRAT peptide, the TDZD-related NP12 and kinase-dead GSK-3 (refs 1–6, 10, 11). These compounds inhibit both GSK-3α and GSK-3β and therefore get around the issue of redundancy. Although off-target effects for these diverse GSK-3 inhibitors could, formally, explain their effects on APP processing, a far more plausible explanation is that these compounds act through GSK-3 inhibition.

Jaworski *et al.*[7] show that APP can be processed in mouse brain in the absence of either *Gsk3a* or neuronal *Gsk3b*, but do not examine redundant functions of *Gsk3a* and *Gsk3b* in APP processing and do not consider the substantial body of work showing that acute inhibition of GSK-3 reduces amyloid-β accumulation *in vivo*. The simplest explanation for these findings[2,5,6] is that both *Gsk3a* and *Gsk3b* can contribute to APP processing and that inhibition of GSK-3 reduces amyloid-β in Alzheimer's disease models.

Christopher J. Phiel[1], Christina A. Wilson[2], Virginia M.-Y. Lee[3] & Peter S. Klein[4]

[1]Center for Molecular and Human Genetics, The Research Institute at Nationwide Children's Hospital, Columbus, Ohio 43205, USA.
[2]Department of Neurology, 3 West Gates Building, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA.
[3]Center for Neurodegenerative Disease Research, Department of Pathology and Laboratory Medicine, 3rd Floor Maloney Building, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA.
[4]Department of Medicine, Division of Hematology–Oncology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104-6148, USA.
e-mail: pklein@mail.med.upenn.edu

1. Phiel, C. J., Wilson, C. A., Lee, V. M. & Klein, P. S. GSK-3α regulates production of Alzheimer's disease amyloid-β peptides. *Nature* **423,** 435–439 (2003).
2. Ryder, J. *et al.* Divergent roles of GSK3 and CDK5 in APP processing. *Biochem. Biophys. Res. Commun.* **312,** 922–929 (2003).
3. Li, B. *et al.* FRAT1 peptide decreases Aβ production in $swAPP_{751}$ cells. *FEBS Lett.* **553,** 347–350 (2003).
4. Su, Y. *et al.* Lithium, a common drug for bipolar disorder treatment, regulates amyloid-β precursor protein processing. *Biochemistry* **43,** 6899–6908 (2004).
5. Huang, H. C. & Klein, P. S. Multiple roles for glycogen synthase kinase-3 as a drug target in Alzheimer's disease. *Curr. Drug Targets* **7,** 1389–1397 (2006).
6. Huang, H. C., O'Brien, W. T. & Klein, P. S. Targeting glycogen synthase kinase-3 in Alzheimer's disease. *Drug Discov. Today Ther. Strateg.* **3,** 613–619 (2006).
7. Jaworski, T. *et al.* GSK-3α/β kinases and amyloid production *in vivo*. *Nature* **480,** doi:10.1038/nature10615 (2011).
8. Doble, B. W. & Woodgett, J. R. GSK-3: tricks of the trade for a multi-tasking kinase. *J. Cell Sci.* **116,** 1175–1186 (2003).
9. Doble, B. W., Patel, S., Wood, G. A., Kockeritz, L. K. & Woodgett, J. R. Functional redundancy of GSK-3α and GSK-3β in Wnt/β-catenin signaling shown by using an allelic series of embryonic stem cell lines. *Dev. Cell* **12,** 957–971 (2007).
10. Sun, X. *et al.* Lithium inhibits amyloid secretion in COS7 cells transfected with amyloid precursor protein C100. *Neurosci. Lett.* **321,** 61–64 (2002).
11. Serenó, L. *et al.* A novel GSK-3β inhibitor reduces Alzheimer's pathology and rescues neuronal loss *in vivo*. *Neurobiol. Dis.* **35,** 359–367 (2009).

doi:10.1038/nature10616

©2011 Macmillan Publishers Limited. All rights reserved



THE IMAGE WORKS / TOPFOTO

PUBLIC HEALTH

Life lessons

The 2009 pandemic arrived suddenly and lethally, exposing our plans to reality. Are we now better prepared?

BY LAURA VARGAS PARADA

Adela Gutiérrez, a door-to-door census-taker for the Mexican government, started feeling unwell with fever, headache and a sore throat, in early April 2009. Gutiérrez, a 39-year-old mother of three, lived in the southern state of Oaxaca. By the time she arrived at the state hospital, she was unable to breathe and was diagnosed with severe atypical pneumonia that might have been complicated by her diabetes. Within four days, Gutiérrez was dead.

Lab tests at the Oaxaca state hospital could not confirm the cause of death, so throat swabs were sent to the National Microbiology Laboratory in Winnipeg, Canada, and to the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia. Ten days later, Cruz would become one of the first confirmed fatalities of a new influenza A (H1N1) virus, which originated in pigs. By then, nearly half of all the people who were in contact with Cruz in the hospital had developed respiratory symptoms and one pregnant nurse had fallen ill. Two months later, the World Health Organization (WHO) declared a global health emergency.

Early observations showed that this strain of flu disproportionately afflicted young people. A study of patients in Mexico found that the median age of fatalities was 39 years (ref. 1). Epidemiologists feared this outbreak could resemble the infamous 1918 influenza pandemic, also an H1N1 strain, which killed an estimated 50 million people (around 3% of the then global population) — half of whom were healthy adults between the ages of 20 years and 40 years old. The US President's Council of Advisors on Science and Technology calculated a possible scenario of 30,000–90,000 deaths in the United States alone. Fortunately, the transmission rate and virulence of the virus could not produce the pandemic that was initially dreaded.

GAINING MORE EXPERIENCE

Nevertheless, the 2009 H1N1 pandemic proved to be a thorough test of the 2005 International Health Regulations (IHR), designed to be the world's first line of defence during public-health emergencies. (IHR is a legal agreement that dates back to 1969 and is binding by 194 party states, including all WHO members.) The 2009 outbreak was the first major assessment of the new IHR and raised critical questions about why it was so difficult to determine the severity of the threat, how effective the preparations were, and whether the crisis was properly managed.

In late 2009, director-general Margaret Chan of the WHO advised a review of the pandemic to learn from any lessons and evaluate whether the IHR fulfilled its purpose. The WHO assigned this task to a newly created IHR review committee, comprised of 25 international experts from diverse scientific fields with experience in public health. The committee presented its final report in May 2011, during the 64th World Health Assembly in Geneva, Switzerland.

The emergence of human infections of avian influenza A (H5N1) in 1997 and the coronavirus causing severe acute respiratory syndrome (SARS) in 2002 expedited international preparations for a pandemic. The IHR was updated in 2005 to respond to these new threats and entered into force worldwide in 2007, and are expected to be fully operational by 2012.

A major point of concern raised during the outbreaks of bird flu and SARS was the performance of disease warning systems in place. To assure a two-way channel of communication between the WHO and the party states, the IHR required each country to establish, by 2012, what it calls a focal point — offices that liaise with the WHO at all times and deploy resources for disease surveillance, early warning systems and the response to a disease outbreak.

The IHR review committee reported the IHR was still "not yet fully operational" worldwide. "There are still many states that have not developed the plans and infrastructure specified in the IHR," says José Ignacio Santos, who heads the infectious disease unit of the school of medicine at the National Autonomous University of Mexico in Mexico City and a member of the IHR review committee. Although three quarters of the 194 party states had plans to coordinate national efforts in the case of a pandemic influenza outbreak, that statistic didn't tell the whole story. Only a dozen of the 128 countries that answered a WHO questionnaire had taken the steps required to put such a plan in practice, including: enacting legislation; allocating funding; putting enough people in place for detection and alert operations; and establishing procedures for surveillance, event detection, risk assessment and information provision[2]. In addition, many National IHR focal points are unable to communicate information about public health emergencies to the WHO in a timely fashion.

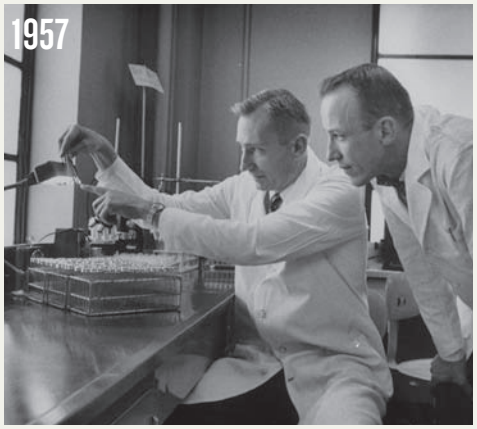
According to the IHR review committee "the IHR played a central role in the global response to the pandemic." In Mexico, the IHR probably saved lives thanks to the progress made towards establishing plans and stocking antiviral medications. But what would happen if a new virus were to appear in countries that are not yet prepared, such as some in Africa and Asia?

The WHO did not escape blame in the IHR review committee's assessment. For instance, the WHO was inconsistent in its definition of 'pandemic'. Some WHO documents described pandemics in terms of the number of deaths and illness caused by the disease, whereas the latest WHO definition is based on the degree



1918
The Spanish flu (H1N1) strikes killing an estimated 50 million people.







Camp Funston, Kansas, site of the first recorded cases of Spanish flu (left), nurse draws water for flu patients (top).

Noted flu researchers. D. E. Rogers and E. D. Kilbourne test for the influenza virus (top), a New York hospital at the height of the epidemic (right).

1957
H2N2 outbreak that began in Asia kills 100,000 people.

1900 1910 1920 1930 1940 1950 1960

of geographical spread. The WHO also failed to come up with consistent measures to assess the severity of an outbreak. The review committee stressed the need for the WHO to strengthen its capacity to mount a sustained response, and to improve its communication policies — in particular by issuing timely guidance in the organization's six official languages and holding routine press conferences. The review committee's report concluded with a stark warning: "The world is ill-prepared to respond to any similarly global, sustained and threatening public-health emergency."

SHARPENING PREDICTIONS

Despite recent efforts to improve preparedness for a global influenza pandemic, there are still many things we don't understand about the virus. "Influenza is the name of one disease but is caused by a family of viruses," says Sylvie Briand, head of the global influenza programme at the WHO in Geneva, Switzerland. "We don't know very much about how these viruses circulate at the global level." It is still an enigma as to when, and to what extent, influenza viruses change their genetic make-up, and how they spread around the globe.

A reminder of how influenza can still surprise us came on 2 September, 2011, when a new swine influenza virus A (H3N2) was isolated from two children in the United States — one in Indiana and one in Pennsylvania[3]. Human H3 and H2 influenza strains were transmitted to pigs in the 1990s. Those viruses evolved in swine and now differ from the seasonal influenza that circulates in humans. People born before the 1990s "still have antibodies to the swine viruses because their immune system was exposed to the original viruses," says Nancy Cox, director of the influenza division at the CDC. "But young kids are very susceptible." The CDC has identified a vaccine candidate against this virus — just in case, says Cox.

In an ominous report published in 2011, researchers studied H2N2 viruses that have not circulated in human populations for several decades, but are still common infections of birds and swine[4]. The study found that people under 50 years of age have little or no immunity to this strain of the virus, reigniting fear that an H2N2 pandemic could, like H1N1, jump from animals to humans.

The 2009 swine flu was a reassortant virus that crossed from pigs to humans (a reassortant virus contains genetic material from two or more related viruses). Cox believes that this exchange between human and porcine viruses is more common than had been previously thought: "This is probably going on in a lot of other countries. In Switzerland they have detected swine infections in humans," Cox says. In a retrospective study published in 2007, researchers found evidence of 50 apparent cases of influenza that were transmitted from pigs to humans. Most of the cases were reported in the United States — probably, Cox says, because the US has a very good surveillance system, but there were also six cases in the Czech Republic, four in the Netherlands, three in Russia, three in Switzerland, and one case each in Canada and Hong Kong[5].

Prior to 2009, three global networks were in place for the early detection and/or surveillance of influenza: the 2005 IHR, the WHO's Global Influenza Surveillance Network (GISN) and systematic event detection (that monitors information published not only formally in journals but also in newspapers and online forums). In March 2009, just weeks before the swine flu outbreak, more than half of WHO member states (104 of 193 countries) had no or very limited seasonal influenza surveillance capacity, according to a 2011 article co-authored by Briand[6]. (Both the United States and Mexico had good surveillance systems, which is why the new virus was identified in those countries early on.) The same report, echoing some of the themes articulated by the IHR review committee report, identified other deficiencies that were revealed during the H1N1 pandemic. Among them were a lack of standards for reporting illness, risk factors and mortality. As a result, different countries used different criteria to define influenza-like illness and collected information on different risk factors. This inconsistency makes it difficult to do the sort of comparisons that would help public health officials assess how well they are doing relative to other places in the world. After the swine flu pandemic, the WHO worked to better define standards and develop a web-based platform called FLuID (Flu Informed Decisions) for direct reporting of epidemiological data, such as intensity of influenza transmission, number of hospitalizations, and risk factors for severe disease. FLuID is similar to the WHO's FluNet surveillance system, which collects information concerning the viral type and subtype.

Scientists are confident that a good surveillance system, along with the GISN monitoring antibodies against the virus, will help detect new viruses before they become a threat to global health. And if this is the case, says Briand, "we have a protocol for rapid containment [if] we detect an outbreak early enough to contain it at its source."

NEW DIRECTIONS

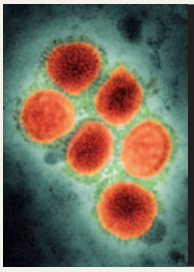
One of the main challenges during an outbreak is to assess the severity of the pandemic disease. Even with today's technology, there are a few critical days of uncertainty after the outbreak is identified, but before enough epidemiological and clinical data are gathered to allow public health officials to plot the most effective response.

"In the 2009 pandemic, some decisions had

FROM LEFT: NATIONAL MUSEUM OF HEALTH AND MEDICINE/SCIENCE PHOTO LIBRARY; NATIONAL ARCHIVES/SCIENCE PHOTO LIBRARY; TIME, BETTMANN/CORBIS & LIFE PICTURES/GETTY IMAGES, BETTMANN/CORBIS, BETTMANN/CORBIS, CDC/SCIENCE PHOTO LIBRARY, KPA/ZUMA/REX FEATURES.

1968
700,000 people die from H3N2 infection (descended from H2N2); originates in Hong Kong.

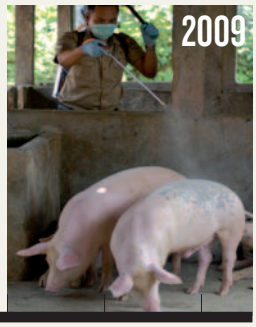




2009
The return of H1N1 — swine flu — surfaces in Mexico and spreads rapidly — including to healthy people, but with a low fatality rate.

A sign in Iowa casts flu as a foreign invader (left).

A pig sty is disinfected in Indonesia (right).



1970 1980 1990 2000 2010

THE NEXT PANDEMIC

Experts recommend the following actions:

- Strengthen surveillance and response in much of the world, especially in under-resourced countries.
- Define pandemic, develop measures to estimate severity.
- Standardize reporting of influenza data.
- Focus on new vaccine production technologies, new antivirals, and extending the shelf lives of medications.
- Vaccine distribution worldwide needs to be revised and strengthened.

to be taken before we had a clear picture," Briand says. Drawing on that experience, she says, health officials are trying to improve their ability to assess an outbreak at its earliest stages. "If we isolate the virus," she explains, "we can determine how different it is from the current circulating viruses and if it is susceptible to available antivirals." Testing blood samples for antibodies to influenza viruses will allow researchers to determine how many people have been exposed to the virus and how prevalence fluctuates over time.

The absence of important data can be overcome. During the 2009 pandemic, determining the fatality rate was sometimes difficult because information about the number of persons infected and the number of deaths was incomplete. Exacerbating the problem, the WHO, as well as national authorities, overwhelmed public health officers with requests for specific data, such as the number of cases. In countries with limited epidemiological and laboratory infrastructure, the need to deliver so many statistics led to an unfortunate diversion of personnel away from patient care. And in countries during the early stages of the pandemic, this information proved to be less useful for assessing severity than rates of hospitalization and complications, and actual number of deaths.

As it turned out, however, other parameters proved to be good proxies to gauge the severity of disease. Measures included the proportion of cases that required hospitalization for treatment or that required intensive care and mechanical ventilation. Other useful data were the proportion of previously healthy individuals without underlying risk factors that developed severe disease. And as always, the severity of a pandemic will depend on conditions inside a country, such as access to health services, general health of the population, and various social and behavioral factors.

As the influenza virus continues to circulate in human and animal populations across the globe, it is crucial to identify changes in the viral genome — especially those mutations that confer resistance to antiviral drugs. In 2009, Briand says, "we were fortunate because the virus was sensitive to neuraminidase inhibitors that we had in our stockpile." Neuraminidase is a viral surface protein that allows the virus to spread from infected cells to other cells (see 'Lines of defence', page S9). But any strategic stockpiling of anti-flu drugs needs to account for the limited shelf lives of the medications. At the WHO, public health officers are looking for technical solutions to prolong the usefulness of the antivirals to at least two or three more years beyond the roughly 5 year shelf lives of today's medications, as well as "encouraging the development of new antivirals", says IHR review committee member Santos.

The best weapon against flu is still the vaccine. Once the H1N1 2009 virus was identified, the WHO needed just one month to select the pandemic virus and develop the seed strains for vaccine development. But that's only the first step. "It takes at least six months to produce significant quantities of a vaccine," Klaus Stöhr of Novartis claimed in 2010 (ref. 7). Stöhr is vice president of influenza strategy at Novartis Vaccines and Diagnostics in Cambridge, Massachusetts. One vexing question is how much vaccine to produce. By the time vaccines against the 2009 H1N1 virus were made available, the production was enough to treat only 10% of the population worldwide. Santos also highlights pharmaceutical deficits. "The technology to produce influenza vaccine has not changed much in the past 60 years," he says.

Even if vaccines are produced in abundance, there's no guarantee that the medication will get to the people who need it. "Vaccine distribution worldwide is the other challenge," says Santos. Access to vaccines, he says, can be improved by legal agreements between manufacturers and countries. Other paths to more equitable access to vaccines include differential pricing so that poorer countries can get the vaccines cheaper — already in practice, but says Santos, needs to be made more universal — as well as vaccine donations by manufacturers to the WHO. At the national level, there is the need to assure the maintenance of the cold chain — the continuous sequence of properly refrigerated vessels to transport the vaccines from factory to the patient — and to plan for local vaccine distribution and administration. "One way to facilitate both vaccine and antiviral distribution," Santos says, "could be to regionalize the effort, giving each WHO regional office the responsibility to secure the resources needed for the area."

No one can predict with precision when and where the next global outbreak will arise. But, in view of what has been learned from recent influenza pandemics, it is possible to define the most likely scenarios and to prepare for them. "It is like your personal life," says Briand. "You cannot predict anything but there are scenarios that are more likely than others." The lessons learned from the latest pandemic "will certainly drive our preparedness to different directions now."

One key lesson is the need to be flexible in our response to unexpected conditions and to be able to respond in times of uncertainty. We must, in other words, be like the flu virus itself. ■

Laura Vargas Parada *is a freelance science writer in Mexico City.*

1. Domínguez-Cherit, G. *et al. JAMA* **302**, 1880–1887 (2009).
2. http://apps.who.int/gb/ebwha/pdf_files/WHA64/A64_10-en.pdf.
3. Nalluswami, K. *et al. MMWR* **60**, 1213–1215 (2011).
4. Nabel, G. J. *et al. Nature* **471**, 157–158 (2011).
5. Myers, K. P. *et al. Clin. Infect. Dis.* **44**, 1084–1088 (2007).
6. Briand, S. et al. *Public Health* **125**, 247–256 (2011).
7. Stohr, K. *Nature* **465**, 161 (2010).



BRAND NEW IMAGE/ GETTY IMAGES

MORBIDITY

A personal response

Some people get horribly sick from the flu, and even die. Others just rest for a few days. What's behind this fateful variation?

BY CHRISTINE JUNGE

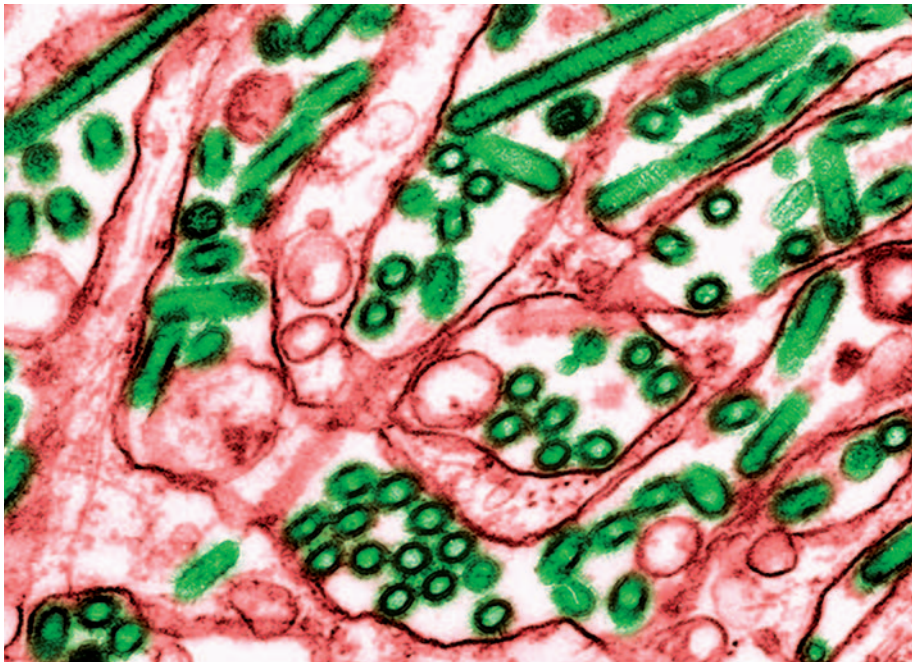
Your spouse had the flu last week and could barely get out of bed for five full days — she was feverish and her aching body exhausted. Then your four-year-old became sick and you had to rush her to the hospital when her fever spiked. Now that you've come down with the virus, you're tired, a bit achy, but that only lasts for two or three days. How is it that the flu virus can have such a variable impact on people?

In order for the flu to pose a danger to any individual, two major factors must align: the host person has to be a clean slate — someone who hasn't been exposed to an influenza virus related to the one circulating; and the virus itself has to be especially potent. The range and severity of symptoms the host experiences depend on the answers to a few other questions, such as whether the person's immune system is compromised, and whether some genetic vulnerability exists.

The first question to delve into when asking about why a particular person gets very sick is: what makes a particular strain of flu more likely to take hold in a human host? The answer lies in the basic structure of the pathogen. The influenza virus always consists of two types of major proteins: one hemagglutinin (H) and one neuraminidase (N). There are at least 16 types of neuraminidase and 9 of hemagglutinin, in principle, capable of forming more than 100 strains of flu virus. In reality, however, the flu viruses that commonly infect humans have been limited to only three types of hemagglutinin and two types of neuraminidase[1], giving six possible flu strains. Each season, there's a chance that, through mutations and recombinations, a novel virus will emerge — one that few, or no, humans have encountered before. Rolling this viral game of dice leads to deadly consequences. 1957 saw the emergence of the H2N2 strain. This represented a major structural shift for the virus; for the previous 40 years, all the flu cases had been caused by H1N1 viruses. This new virus killed 70,000 people in the United States alone[2].

A more recent example of a novel virus was the H5N1 swine flu that circulated in 2009. Unlike seasonal flus, which are most severe in the elderly and those with other health problems, this virus killed a large number of young people — about a quarter of the deaths were in people younger than 24 years (many of them without pre-existing illness), and more than 60% were in people aged 25–64 years — again, many of them otherwise healthy. The reason for this reversal of fortunes was that the virus was similar to ones that had circulated many years ago. Many elderly people, therefore, had some immunity to it, but their children and grandchildren had never been exposed to any influenza like this one and so lacked any resistance to it[3].

Another issue that sets flu apart from some other infectious diseases is that antibodies against influenza don't have a very long shelf



CDC/SCIENCE PHOTO LIBRARY

Up close and personal: H5N1 (green) in cell culture.

life. In certain groups of people, there are additional concerns: after peaking, antibody levels decline faster in older people than in younger people. In people with compromised immune systems, such as someone with HIV/AIDS or someone taking certain cancer drugs, it's double trouble — antibody levels don't rise as high after exposure to the virus in the first place, nor do they last as long. "It's difficult to say whether any individual is still protected after a year or not, which is why we recommend getting the flu shot each year," says Lisa Grohskopf an infectious disease specialist who works in the influenza division of the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia. She adds that protection can wane from the beginning of one flu season to its end, since the season can stretch for seven or eight months.

Infants or young children who haven't come into contact with the flu virus before — either by getting infected or from vaccination — don't have antibodies specific to the viruses circulating in a given year, putting them at risk for getting the flu.

For reasons that aren't always clear, according to microbiologist Peter Palese at the Mount Sinai School of Medicine in New York, people with certain chronic illnesses also have a harder time resisting flu or fighting it once they do get it. Flu is more dangerous for people with heart disease, diabetes, asthma, chronic obstructive pulmonary disease (COPD) and morbid obesity. Palese explains that some of those diseases, such as asthma and COPD, cause persistent lung problems, so any illness that causes fluid to accumulate in the lungs, as flu does, would be more problematic for these people. For the obese, he says, breathing can be difficult, so the act of clearing mucus from the lungs might also be difficult, and the virus would linger in the lungs longer. Grohskopf adds that various health problems can create a vicious cycle. For instance, she says, "diabetes affects the immune system and so may affect the body's ability to fight influenza; in turn, being ill with the flu can cause difficulty with keeping blood sugar under control," an outcome that can complicate diabetes.

Flu's effects can be amplified not only by illness but by our genes. Indeed, just as the genetics of the flu virus itself (which determines which H and N type it is) can make it stronger or weaker, the genetics of the person who encounters the influenza virus can make that person more or less likely to become ill. One thing that has become clear in recent years is that single nucleotide polymorphisms in one of a number of genes governing immunity affects a person's susceptibility to infectious diseases, and how sick he or she becomes from those infections. A 2008 study by Ute Vollmer-Conna, a geneticist at the University of New South Wales in Australia, who studies the genetic response to infection, advanced that idea. Vollmer-Conna's research looked at 300 people infected with one of three viruses: Epstein-Barr, *Coxiella burnetii* (which causes Q fever) and the Ross River virus[4]. Her conclusion: those who had the most severe symptoms, and those who were ill for the longest, had polymorphisms that significantly influenced pro- and anti-inflammatory cytokine production. Vollmer-Conna explains that these genetic predispositions intensify the inflammatory response to an acute infection, as well as prolonging recovery time. Although the scope of her study did not include influenza, she hypothesizes that "the same principles apply across a much broader range of common infections, including flu".

> ***"The exuberance of the inflammatory response is likely at the root of why we have symptoms of infection."***

Another, more recent study looked at the changes in gene expression caused by the flu virus specifically. Researchers exposed 17 healthy subjects to flu virus type H3N2. Nine came down with flu symptoms and eight did not. When the scientists looked at blood from each subject, they noticed that different genes were up-regulated, depending on whether the person fell ill or not. "Individuals who became symptomatic had up-regulation of genes involved in the inflammatory response," says Geoffrey Ginsburg, director of the center for genomic medicine at the Duke Institute for Genome Sciences and Policy in Durham, North Carolina, and one of the authors of the study. "The exuberance of the inflammatory response is likely at the root of why we have symptoms of infection," he says. In those who did not become ill, he explains, the genes that were up-regulated were those involved in protein synthesis and oxidative stress. "Asymptomatic subjects may have mechanisms at work that prevent active infection and therefore reduce the inflammatory response."

Aside from more research into the human genes that control the immune response to flu, there are plenty of other lines of investigation that would help explain why different people respond to flu so differently. Vollmer-Conna argues for more research to tease out which genes and biological processes are responsible for recovery after an infection. This knowledge, she writes, could let doctors identify those at risk for severe illness following infection with common diseases, and therefore set them up with special prevention and treatment programmes.

In a 2004 research paper, Palese posed several questions, including what makes it possible for some influenza viruses to hop from one animal species to another; questions that he hoped researchers would have answered by now. He concedes that seven years later, the questions on his list are still pertinent. One in particular that still remains an important area of research, says Palese, are the viral genes that determine transmission versus the host or environmental factors that promote its spread.

Clearly researchers have made some progress, but as the poet Robert Frost once wrote, there are miles to go before we sleep. Frost was writing about stopping in the woods on a snowy evening in winter — which, of course, is flu season. ■

Christine Junge *is a science writer in North Easton, Massachusetts.*

1. Palese, P. *Nature Medicine* **10**, S82–S87 (2004).
2. Dolin, R. *How to understand your risk and protect your health (Harvard Health Publications*, 2009).
3. Vollmer-Conna, U. *et al. Clinical Infect. Dis.* **47**, 1418–1425 (2008).
4. Huang, Y. *et al. PLoS Genet.* **7**, e1002234 (2011).

nature OUTLOOK

INFLUENZA

8 December 2011 / Vol 480 / Issue No. 7376



COVER ART: NIK SPENCER

Editorial
Herb Brody, Michelle Grayson, Tony Scully

Art & Design
Wes Fernandes, Alisdair MacDonald, Andrea Duffy

Production
Karl Smart, Emilia Orviss, Leonora Dawson-Bowling

Sponsorship
Patrick Murphy, Reya Silao, Yvette Smith, Gerard Preston

Marketing
Elena Woodstock, Hannah Phipps

Project Manager
Helen Anthony, Claudia Deasy

Art Director
Kelly Buckheit Krause

Magazine Editor
Tim Appenzeller

Editor-in-Chief
Philip Campbell

Amidst the series of tragic events that was the first half of the twentieth century, the 1918 influenza pandemic stands out for its swift lethality. Over the course of about 18 months at the close of the First World War, an estimated 50 million people died from this viral infection.

Since the last flu pandemic in 2009, public health officials have started to stand guard against another catastrophe (page S11). Yet echoes of this outbreak have reverberated through the decades and the disease, in its various guises, kills at least a quarter of a million people each year worldwide. The virus is adept at adaptation, leaving scientists chasing a moving target. The virus's surface proteins mutate rapidly and can combine into dozens of variants (page S2); the antibodies we produce to fight one year's flu strain can't stop the others.

Influenza evades the effects of drugs developed against it. First-generation antivirals are now almost useless, and newer classes of drugs — the neuraminidase inhibitors oseltamivir (Tamiflu) and zanamivir (Relenza) — are beset by resistance(page S9). And, owing to a number of factors, including genetics, individual responses to the virus vary widely; researchers are starting to get a handle on why some people infected with influenza become very sick or die while others are unscathed (page S14).

Fears of widespread contagion have grown because the virus seems to be making a habit of jumping from its original bird hosts into domesticated mammals and birds (page S4). Perhaps the great hope lies in flu prevention. Researchers are reporting progress in developing a vaccine that could be effective against a broad range of flu strains (page S6), a feat that might mark a turning point against this killer.

We are pleased to acknowledge the financial support of Sanofi Pasteur, Crucell and Baxter and we also acknowledge the unrestricted grant from F.Hoffmann-La Roche Ltd. As always, *Nature* retains sole responsibility for all editorial content.

Herb Brody
Supplements Editor, Nature Outlook.

CONTENTS

S2 EPIDEMIOLOGY
Racing against the flu
Life isn't easy living with one of humanities biggest killers

S4 Q&A
The flu catcher
Richard Webby's keeps track of the various ecological niches of influenza circulating in animal populations

S6 PREVENTION
Vaccine for all seasons
Targeting the more stable parts of influenza's surface proteins is showing some promise, and might lead to a universal vaccine

S9 DRUGS
Lines of defence
In search of other antiviral treatments to prepare for the next pandemic

S11 PUBLIC HEALTH
Life lessons
Without learning from influenza's tragic history we are doomed to repeat it

S14 MORBIDITY
A personal response
Some people get sicker than most, and others die. What's behind these drastic differences?

COLLECTION

S17 **Live attenuated influenza virus vaccines by computer-aided rational design**
Steffen Mueller et al.

S21 **Long-term evolution and transmission dynamics of swine influenza A virus**
Dhanasekaran Vijaykrishna et al.

S25 **Dissolving polymer microneedle patches for influenza vaccination**
Sean Sullivan et al.

Nature Outlooks are sponsored supplements that aim to stimulate interest and debate around a subject of interest to the sponsor, while satisfying the editorial values of *Nature* and our readers' expectations. The boundaries of sponsor involvement are clearly delineated in the *Nature Outlook* Editorial guidelines available at http://www.nature.com/advertising/resources/pdf/outlook_guidelines.pdf

CITING THE OUTLOOK
Cite as a supplement to *Nature*, for example, Nature **Vol XXX**, No. XXXX Suppl, Sxx–Sxx (2011). To cite previously published articles from the collection, please use the original citation, which can be found at the start of each article.

VISIT THE OUTLOOK ONLINE
The *Nature Outlook Influenza* supplement can be found at http://www.nature.com/nature/outlook/Influenza/

All featured articles will be freely available for 6 months.

SUBSCRIPTIONS AND CUSTOMER SERVICES
For UK/Europe (excluding Japan):Nature Publishing Group, Subscriptions, Brunel Road, Basingstoke, Hants, RG21 6XS, UK. Tel: +44 (0) 1256 329242. Subscriptions and customer services for Americas – including Canada, Latin America and the Caribbean: Nature Publishing Group, 75 Varick St, 9th floor, New York, NY 10013-1917, USA. Tel: +1 866 363 7860 (US/Canada) or +1 212 726 9223 (outside US/Canada). Japan/China/Korea:Nature Publishing Group — Asia-Pacific, Chiyoda Building 5-6th Floor, 2-37 Ichigaya Tamachi, Shinjuku-ku, Tokyo, 162-0843, Japan. Tel: +81 3 3267 8751.

CUSTOMER SERVICES
Feedback@nature.com

Copyright © 2011 Nature Publishing Group

EPIDEMIOLOGY

Racing against the flu

Influenza mutates fast and spreads easily, earning a place among humanity's biggest killers.

BY DUNCAN GRAHAM-ROWE

Influenza kills more than 250,000 people each year, worldwide. And with new strains ever emerging, there is always the threat of a repeat of the Spanish flu pandemic of 1918, which claimed more than 50 million lives. Indeed, despite vaccines and antiviral drugs that offer some protection to the vulnerable, nearly a century after that catastrophe we still seem no closer to being able to predict and ultimately stave off such a virulent outbreak.

One reason this has proved so difficult is the speed at which this acute viral infection can spread, transmitted by coughs and sneezes. "It's a very stable respiratory virus that survives well on surfaces and people's hands," says flu specialist Anthony Mounts with the Global Influenza Programme at the World Health Organization in Geneva, Switzerland.

Moreover, the flu virus changes all the time, mutating at the genetic level, which alters the surface level — in the hemagglutinin and neuraminidase antigens that coat it (and provide the familiar H-N designations). These proteins are what our antibodies attach to; by continually changing, these antigens enable the virus to infect people who have developed immunity to other strains. It is because of its ability to rapidly mutate that the virus is able to develop resistance to previously effective drugs, such as oseltamivir (marketed as Tamiflu by Switzerland-based Roche) and amantadine.

On average between 5–15% of the global population are infected each year. Most at risk are children younger than 2 years and adults 65 years or older, as well as people of any age with asthma, diabetes or chronic heart disease. But no one is entirely free from risk. "This was particularly notable in this last pandemic," says Mounts, referring to the 2009 outbreak, in which as many as half of those contracting the disease had fully functioning immune systems and were perfectly healthy beforehand.

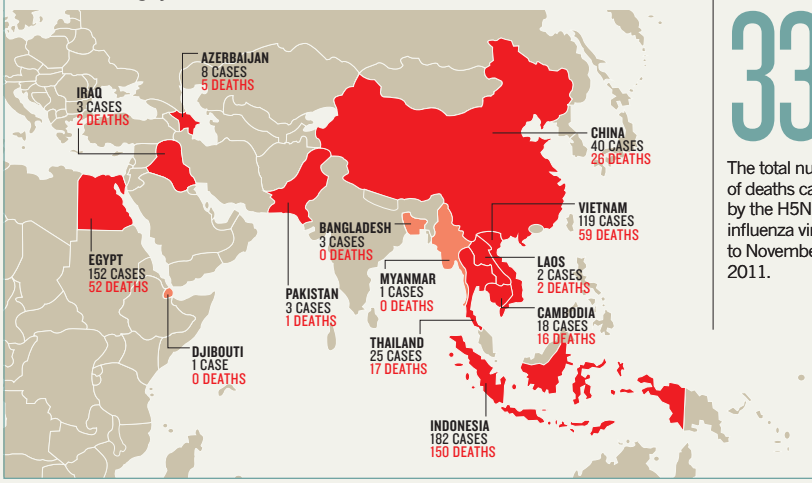
NATURE.COM
Keep up to date with the latest news and research on influenza:
go.nature.com/azxjrn

THE FAST OR THE FURIOUS

To exact the greatest toll, influenza needs to be both fast-spreading and highly lethal. Unlike the 1918 'Spanish flu', the two recent outbreaks — H5N1 'bird flu' and H1N1 'swine flu' — only had one attribute each.

THE SLOW BUT DEADLY SPREAD OF H5N1

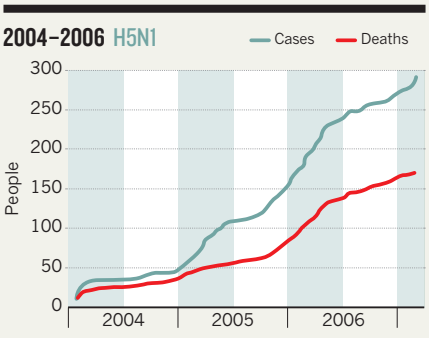
Between 2003 and 2011, H5N1, commonly known as bird flu, spreads across Asia but remains largely an avian virus.



335
The total number of deaths caused by the H5N1 influenza virus up to November 2011.

CASES ON THE RISE

The mortality rate has risen with the number of cases worldwide



60% approximate proportion of fatalities for immunocompetent people infected with H5N1.

H5N1 2003

1890 — RUSSIA
First recorded influenza pandemic. 1 million die worldwide.

1918 — H1N1
PANDEMIC
The origin of H1N1 has been the subject of many studies — some suggest it may have gone directly from birds to humans, while others suggest it may have involved incubation in an intermediate host, such as the pig.

50M
The 'Spanish Flu' caused by the H1N1 influenza virus, kills approximately 50 million people.

1957 — H2N2
1959 — H7N7
ASIA
The second influenza pandemic strikes in Asia, killing 100,000 people worldwide.

Several new strains of influenza emerge throughout the twentieth century, circulating without turning pandemic.

1968 — H3N2
HONG KONG
Flu pandemic kills 700,000 people worldwide. Virus still circulates today.

1977
H1N1 outbreak of Russian flu in China, similar to strain in circulation prior to 1957, spreads around the world disproportionately infecting people under the age of 23.



The virulence of an outbreak is a combination of the viral lethality and infectiousness. These, in turn, depend on which of the three varieties of human flu virus is the pathogen: type A, type B or type C. Type A is by far the most virulent; with 16 known subtypes, it mutates into countless different genetic and antigenic strains at a rate far faster than type B or type C. This ability to mutate quickly allows type A flu to spread faster. All the flu strains discussed in this *Nature Outlook* are influenza type A.

Some of the most virulent strains encountered are those that have managed to pass from other species to humans. The avian H5N1 flu virus had a very high lethality, killing 56% of those who were infected. However it was also not terribly infectious: between 2003 and 2006 there were only 263 cases of H5N1, mitigating its overall impact. In contrast, the H1N1 swine flu virus of 2009–2010 had a case fatality rate of only around 0.03% but, as it can be passed easily from human to human, it was responsible for more than 13,000 cases in a single month.

INFLUENZA MENAGERIE

Flu has been found in a variety of mammals, from horses, cats, dogs, and pigs, to seals, ferrets, camels and even whales. But it is mainly through wild aquatic birds that the virus is able to jump species — most likely because the virus originated in these avian species. Although it is next to impossible to predict where and when an inter-species jump is likely to take place, or which of these will result in a particularly virulent human strain, clues are beginning to emerge.

By analysing the genetic and antigenic variations of 13,000 strains of the H3N2 virus, a group led by infectious-disease experts Colin Russell and Derek Smith at the University of Cambridge in the UK were able to track the evolution of the virus as it moved around the world between 2002 and 2007. They found that, at least for the common H3N2 virus, strains tend to originate in East and Southeast Asia, and mutate their way around the world until they end up in what Smith calls an "evolutionary graveyard" in South America — a continent that, in terms of flu at least, he says, is therefore "the safest place to be because you get the biggest warning."

The conditions in Asia are ripe for new influenza strains to emerge, says Russell. For one thing, the virus is able to exist in circulation almost perpetually, giving it more of a chance to mutate. Because tropical countries have no winter, seasonal influenza epidemics typically occur during the rainy season, which can happen at different times of the year in neighbouring places like Bangkok and Kuala Lumpar, says Russell. This means that within East and Southeast Asia there will be an influenza epidemic happening somewhere at any point in a year. "Given the travel and trade between cities and countries in the regions, influenza seems to be able to spread readily from place to place," says Russell. "We often talk about it as being similar to runners passing a baton in a relay race, as viruses move from epidemic to epidemic."

It's a race that researchers, public health authorities, and the public continue — with mixed success — to try to disrupt. ■

Duncan Graham-Rowe *is a science writer in Brighton, UK.*

0.03%

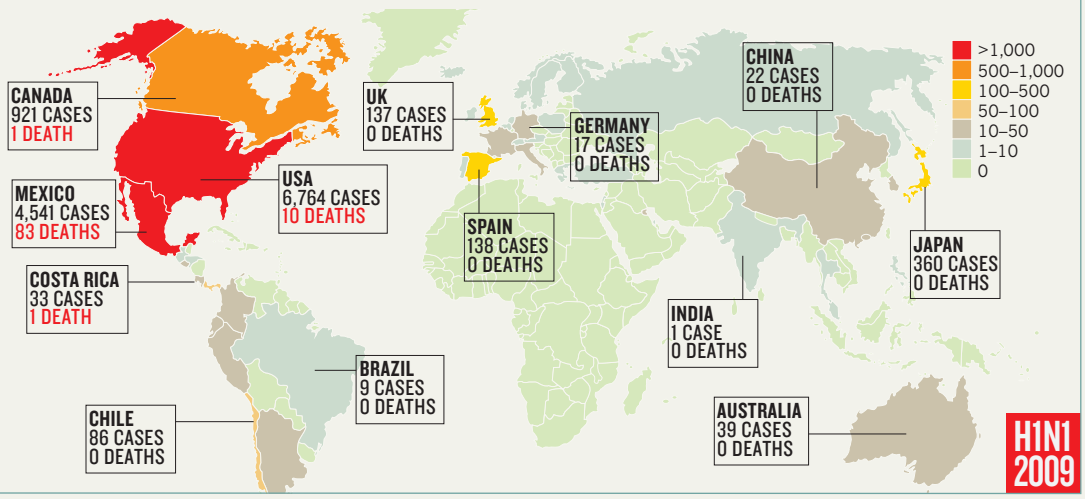
approximate proportion of fatalities for immunocompetent people infected with H1N1.

18,449

Confirmed deaths in 214 countries caused by the H1N1 virus up to August 2010.

H1N1, SNAPSHOT OF A RAPID PANDEMIC

In May 2009, 48 countries reported 13,398 cases of influenza H1N1, the first influenza pandemic in four decades.



71–167 BILLION

The estimated cost (US$) of seasonal flu to the US economy each year

250,000+

People worldwide die each year from seasonal influenza infection.

1997: First case of direct transmission from birds to humans.

1999: FDA approves the neuraminidase inhibitor oseltamivir.

ASIA
H5N1, first detected in humans in 1997, evolved in Hong Kong chickens, remains largely an avian virus but known to kill people in close contact with birds (see map, left).

2003–2004
Outbreak of the avian influenza H7N7 in Holland, among mainly poultry workers.
Panama strain of H3N2 mutates into Fijian strain of H3N2.

MEXICO
First pandemic of influenza in 40 years. Labs in the US confirm the re-emergence of the dreaded H1N1 strain, the virus behind the 1918 pandemic.

2009: Cases of Tamiflu resistant H1N1 reported.

2010: August, the WHO announces end of H1N1 pandemic.



SOURCES: WHO/GIP AND FAUCI, A. S. MILBANK MEMORIAL LECTURE 1–22 (2006).



ST. JUDE PHOTOGRAPHY

Q&A

The flu catcher

Richard Webby studies the ecology of influenza, trying to better understand how certain strains of influenza can leap across the species divide from animals to people. Nature Outlook *sat down with him to learn more about his research.*

What is the basic ecology of influenza?
All influenza type A viruses likely originated in wild aquatic birds, and they've moved out to other hosts from there. The next layer is the intermediate hosts. Domestic animal species — mainly poultry and pigs — most often transmit the viruses to humans. Once the viruses have found their way from their natural wild hosts into domestic species, they behave differently. Something about these domestic reservoirs seems to increase the virus's ability to infect people.

What are the biggest unanswered questions when it comes to these interspecies jumps?
What, all of a sudden, makes an animal virus a human virus? What specific virologic factors, what host factors, what environmental factors, align to allow that to happen? We're starting to gain greater understanding of the types of viruses circulating in animal populations, but we don't have a firm grasp on which of them we should be most concerned about. The wildlife component is probably still one of the biggest missing pieces and it's one of the hardest to study. At least with the domesticated species you don't have to chase them around too long to catch them.

"If there's anything that keeps many of us up at night, it's the H5 virus."

Still, we have learned a lot in the past decade. You pick up an old textbook and it'll say if a virus is to jump from the aquatic bird reservoirs to humans, it probably has to go through an intermediate mammalian host, like a pig. The H5N1 strain of the influenza A virus— the bird flu that emerged in Asia in the late 1990s — taught us that viruses can also use domestic poultry as the intermediate host. I can't think of many viruses that can pass directly between birds and humans besides influenza — certainly none that have the same capacity to cause human disease.

Did bird flu teach us anything else?
H5N1 resulted in a lot more resources being put into understanding the animal reservoirs of influenza. Before bird flu emerged, what information we had typically came from fairly confined geographical regions. But in the last ten years we've increased that knowledge a lot, and now we understand more about the virus in different parts of the world. Of course, there are still many gaps to fill, but we're certainly much better off.

Do wild birds experience flu like people do?
Not really. Flu-infected wild birds don't usually appear to be sick. It's a fairly stable host–microorganism relationship, most likely evolved over millions of years, and the virus doesn't cause many symptoms in the birds. When influenza viruses first enter a new host species, they might cause a lot of disease but mortality drops over time. The virus is now simply using wild birds as a means to propagate and maintain itself. Highly

pathogenic forms of avian flu typically arise through evolution in domestic poultry species, and when they transmit back to wild birds, sometimes they've become more virulent or severe and do cause mortality in that bird species. But that's the exception rather than the rule.

How do wild birds transmit flu amongst themselves?

That's still unclear. One research angle is trying to understand hotspots of virus activity. The Delaware Bay region on the US east coast, which my colleagues and I have been studying, is a prime example. Shorebirds stopping in on their northward migration in May to feed on the eggs of spawning horseshoe crabs seem to have a very high prevalence of influenza. There's something special about that environment — probably that there are huge numbers of birds, many of them naive for influenza, coming together in one location at one time — and the virus spreads very rapidly. A few studies looking at those same bird populations further up or down their migratory path have shown much lower influenza burdens. Flu hotspots also appear in Australia, and we have some ongoing studies in Alberta, Canada. But they probably occur in many other parts of the world as well, where waterfowl gather to feed or breed.

How many types of influenza viruses circulate in wildlife?

The influenza virus has two major proteins on its surface: hemagglutinin (H) and neuraminidase (N). Within the aquatic bird populations there are 16 serologically and genetically distinguishable hemagglutinins and 9 different neuraminidases. This is where the flu subtypes get their names, like H5N1. Most — but not all — of the 144 possible combinations have been found circulating in wild birds. When we start to move away from that natural reservoir into domestic animals and humans, the diversity of viruses diminishes.

Are other animals besides wild birds and domesticated poultry and pigs part of the influenza ecosystem?

In terms of wildlife, it's not just birds. We also have good serologic evidence that wild pigs can be infected by flu, and there are very likely other wild hosts that we know nothing about at all. When it comes to domesticated species, horses tend to harbour a horse-specific lineage of influenza virus, though that virus recently transferred to dogs. And the H5N1 bird flu virus and the 2009 H1N1 pandemic virus, which was widely called swine flu, have both jumped into dogs and cats. Exactly what role those animals play in terms of the ecology of flu is unclear, and I don't think we have a good handle on what the true prevalence of the virus is within their populations. But certainly they can be hosts, they can get infected from other species, and they can probably transfer the virus within their own species as well.

➲ NATURE.COM
Scientific American looks at where flu is lurking:
go.nature.com/lqkgwr

Why does the influenza virus seem to be so active in jumping between animals and people right now?

That's the question everyone is asking. Certainly the much larger demand on protein from the global population and rearing animals in larger numbers in smaller areas play a role in the evolution of some of these viruses. And these practices are also bringing domestic species into more contact with wild species.

What questions are you trying to answer with your current research?

We're focused on understanding how viruses behave over time within a given animal population, whether that be wild bird populations, poultry markets or swine herds. We are going back to some locations repeatedly to understand how the viruses themselves and their prevalences change over time. Then, if we do detect interspecies transmissions, we study these viruses in animal models to understand the virologic basis for the differences in epidemiology that we see on the ground. For instance, we're undertaking a fairly large study with colleagues at the National Research Center in Cairo to study people regularly in contact with poultry to understand the real prevalence of the H5N1 bird flu virus jumping into humans. Typically, we only detect the virus in people with severe disease symptoms, since they're the ones that come into hospitals. But that's very likely just the tip of the iceberg — and we have absolutely no idea how big the iceberg is. We've also been going back to some US swine farms to study circulating viruses — how they come into these populations, how long they stay, and how they leave.

"The wildlife component is probably still one of the biggest missing pieces and it's one of the hardest to study."

It's very practical research. For example, here at this WHO meeting we're using detection data from over the past year to make sure that we have the right diagnostic capacities to detect the most current strains, and ultimately have the right vaccines to protect the public.

Nearly 3 years on, what's the lesson from the 2009 H1N1 swine flu pandemic?

The most important lingering question is finding out what was special about the H1N1 pandemic virus that allowed it to become established in humans. There's a lot of work studying the differences between it and the viruses that were circulating in pigs before the pandemic that never successfully jumped to humans. Had somebody identified the pandemic virus in pigs before the outbreak, nobody would have been jumping up and down. It had none of the hallmarks that would have had us scuttling to make vaccines against it — no functional PB1-F2 protein, which contributes to a flu virus's lethality, for instance, and no novel gene segments. That's obviously not good. We hope this research will help us predict which viruses we need to be most concerned about.

The H5N1 bird flu virus so far hasn't shown much ability to pass from person to person. What's the likelihood that the worst-case bird flu pandemic scenario will happen?

That's a tough one. The threat of H5N1 becoming a pandemic is likely lower than for any H1, H2 or H3 influenza virus because H5 isn't as contagious in humans. But if there is a virus we do not want in humans, it is the H5 because it's so deadly. One of the hallmarks of highly pathogenic influenza viruses is that they accumulate additional amino acids in the hemagglutinin protein, which are thought to allow the virus to generate a systemic infection rather than just a respiratory tract infection. Only H5 and H7 viruses are known to do this naturally.

H5N1 has had 10 years to try, which tells us that it's not one, two, or even three or four changes that are required for it to become a human pathogen — it's probably more. On the other hand, there are examples of viruses taking several decades to jump species, despite prolonged contact. So we now have a little bit of confidence that H5N1 can't become transmissible easily, but unfortunately I think it's still a possibility. If there's anything that keeps many of us up at night, it's the H5 virus.

Are there other influenza strains that researchers have their eye on as well?

There are a handful of viruses that likely pose the most threat. H5 being one because we know it infects humans and when it does, it causes severe disease. H7 poses a similar concern. Then there are the H2s, which have the capacity to be a successful human pathogen. There was an H2 pandemic in 1957, and the virus disappeared in 1968, so there's a large percentage of the population that has no immunity to that virus now.

What do you find most interesting about influenza?

It still amazes me how little we know about these viruses. There's not a whole lot to them—they're fairly simple viruses — but we really have no good feel at all for what allows them to jump species. It's an exciting field to be in. ■

Richard Webby, a virologist at St Jude Children's Research Hospital in Memphis, Tennessee, directs the World Health Organization's collaborating center for studies on the ecology of influenza in lower animals and birds.

Interview by **Rebecca Kessler**, *a freelance science journalist in Providence, Rhode Island.*



DUNG VO TRUNG/LOOK AT SCIENCES/SCIENCE PHOTO LIBRARY

Scientists developing a human vaccine against the avian influenza.

PREVENTION

Vaccine for all seasons

As researchers map the stable parts of the proteins that stud the surface of influenza, the decades-long quest for a universal flu vaccine is showing signs of progress.

BY JANA SCHLÜTTER

While the eyes of flu researchers were focused on Asian bird markets in 2009, a pandemic began to emerge at the opposite end of the world in Mexico and California: H1N1 influenza, a version of the 1918 flu virus circulating in pigs for nearly a century, had suddenly leaped back into the human population, which now lacked herd immunity. The virus took the world by surprise.

It took months to develop a vaccine matched to H1N1— too long to thwart the pandemic, which peaked in March 2009 and then again in early November, 2009. "We were fortunate that the virus was not really a nasty bastard and did not kill so many people," says virologist Robert Webster, whose laboratory at St Jude Children's Research Hospital in Memphis, Tennessee, tracks flu viruses and guides the development of flu vaccines to stop them. "If we had an H5N1 virus that was transmitting from human to human: God help us! We would have run out of antivirals almost overnight."

The 2009 outbreak reignited the hunt for a universal vaccine. In contrast to the seasonal flu jab or infection, the spreading pandemic virus, because it was so unlike its predecessors, elicited more cross-reactive, broadly neutralizing antibodies than usual — and now scientists had the technology to find them. One patient even managed to make an antibody that could inactivate all 16 subtypes of the influenza type A virus — the so-called FI6 antibody. (Type A influenza is the most virulent kind, and is responsible for virtually all major human flu outbreaks.)

➲ NATURE.COM
For some of the latest research on influenza vaccines:
go.nature.com/tezyql

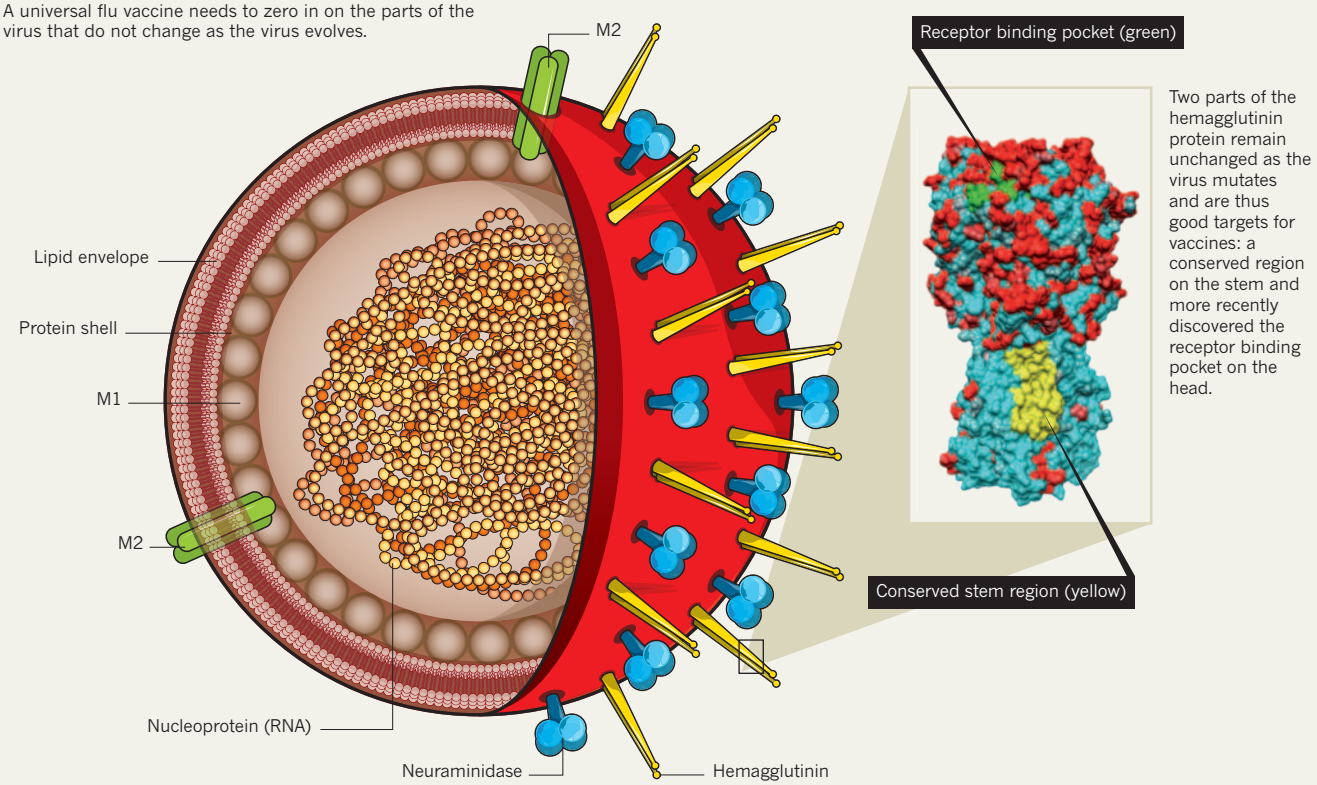
It was, however, a rare find. Antonio Lanzavecchia, an immunologist at the Institute for Research in Biomedicine in Bellinzona, Switzerland, and his colleagues had to screen 104,000 white blood cells (B cells) collected from eight donors until they found one cell that produced the FI6 antibody. FI6 binds to hemagglutinins representative of all the 16 subtypes of type A influenza, including the H1N1 swine flu and avian influenza H5N1, as well as to the more common H3N2 strains. In theory, this FI6 antibody provides a blueprint for designing a one-size-fits-all shot. "People have been searching for it forever," says Webster. "Now, it looks as if the Holy Grail for flu really might be achievable."

STEADY TARGET

The flu virus can skillfully play a chess-like game with the human immune system. Its surface is packed with varieties of two main proteins: neuraminidase and hemagglutinin. Hemagglutinin in particular plays two vital roles: its stem contains the machinery that allows the virus

TARGETING A SHAPE-SHIFTING VIRUS

A universal flu vaccine needs to zero in on the parts of the virus that do not change as the virus evolves.



Two parts of the hemagglutinin protein remain unchanged as the virus mutates and are thus good targets for vaccines: a conserved region on the stem and more recently discovered the receptor binding pocket on the head.

to fuse with host cells; while flashy loops on its globular head act as decoys. To an antibody, each virus appears to be covered with something resembling a dense tropical forest. The freely moving chains of amino acids through the treetops offer alluring binding sites. But these chains mutate and alter their shapes rapidly. The antibody that the host's immune system tailors to fit into the branches of one virus will be limited to that virus, and not work against any new influenza.

Because existing flu vaccines mimic a natural infection, they fall into the same trap: even if our body manages to make a few broadly neutralizing antibodies they are vastly outnumbered by strain-specific ones that bind to the globular head — and most people have to get a flu shot each year to regain immunity.

"When we understand the basics behind this biology we will be in the driver's seat to design a lot of vaccines."

To sidestep the trap, the immune system would need to refocus its attack, perhaps onto the bases of the trees: a highly stable region of the hemagglutinin stem that tends to stay the same even after the virus mutates. FI6 and most other cross-reactive antibodies push their way in and go for these hard-to-find targets. "This is the first time that an antibody can neutralize every single influenza virus," says Rino Rappuoli, head of vaccines research at Swiss pharmaceutical company Novartis. But while calling FI6 a "milestone", Rappuoli is cautious about seeing it as the key to making a universal vaccine. "That's tough," he says. "We don't know how to do it yet, but it is a dream that can come true."

ROOT OF THE PROBLEM

There are two ways to use broadly reactive antibodies: as a passive immunization to treat severely ill patients who did not respond to antivirals, or as a template to develop a prophylactic vaccine for everyone. The latter is the more difficult task — it requires sophisticated engineering of the antibody to derive a protein that not only fits the antibody's binding site but also manages to coax the human immune system into producing enough ammunition to prevent infection. Many researchers favour starting with the first option. "It's effective, it's extremely broad, and it's human — there should be no problem injecting antibodies into humans," says Lanzavecchia. "To be frank, our first goal is to use the antibody as a therapeutic."

Gary Nabel, director of the Vaccine Research Center at the National Institute of Allergy and Infectious Diseases (NIAID) in Bethesda, Maryland, agrees on the difficulty in making a universal preventative flu vaccine. "We've been doing similar work with HIV," he says. Creating a prophylactic vaccine "sounds simple, but it takes time." The challenge, then, is in creating a vaccine that will induce this specific yet unnatural response. Even if one presents a specific site of a protein to the immune system, there are many ways in which antibodies can bind to it. "That's clearly an important problem to solve for any infectious disease, and flu can be our poster child," says Nabel. "When we understand the basics behind this biology we will be in the driver's seat to design a lot of vaccines that could be very, very promising."

While eliciting 'anti-stem' antibodies in humans is difficult, it is not impossible. Nabel and his team were the first to apply a so-called prime-boost approach in humans to achieve this goal. This technique entails giving a combination of two vaccines administered in sequence to induce the strongest immune response possible against a broad array of influenza strains, including H5N1, H5N2 and H9N2. The priming shot was a DNA vaccine: small, circular pieces of bacterial DNA genetically engineered to code for a specific protein that targets an avian flu hemagglutinin. Then, 24-weeks later, volunteers were given a booster vaccine made of whole inactivated H5N1 virus. This regimen enhanced the immune response against avian flu. In addition, three individuals were able to make broadly neutralizing antibodies directed at the hemagglutinin stem, a result that Nabel says is a "proof of concept" for a universal influenza vaccine.

The mice in microbiologist Peter Palese's lab at Mount Sinai School of Medicine in New York are already protected against various flu strains

by experimental, prophylactic flu vaccines. Palese has shown that two truncated versions of hemagglutinin, one with its globular head chopped off and the other only partly expressed, work better than full-length hemagglutinin, but they're not perfect: they offer good protection against H3 strains of influenza virus, but they are weaker against H1. Moreover, Palese echoes the familiar caution from anyone working with animal models: "It's promising," he says, "but mice are not men".

Nonetheless, scientists and vaccine companies around the world are hard at work. Advances in high-throughput technology allow them to screen individual B cells from healthy volunteers or patients and search for rare antibodies. In 2008, a group based at the Dutch company Crucell was the first to find broadly neutralizing antibodies against influenza in humans. Apart from exploring therapeutic approaches, they are also trying to vaccinate rodents with conserved viral structures.

Normally, the immune system overlooks the virus's stem (because it's a target that's hard to see and to get to) and produces instead many antibodies that target the viral protein's head. Most broadly neutralizing antibodies, on the other hand, bind to conserved domains on the stem. Crucell is trying to recreate these conserved areas and expose them to the immune system to induce the production of broadly neutralizing antibodies. It's crucial to use the right binding domain, because otherwise the antibody might be just as strain-specific as anything that binds to the head. "We know that we can rebuild the stem in a way that it looks like the binding domain," explains Crucell immunologist Katarina Radosevic. But the question is how to rebuild these structures in a stable form so that they can induce an immune response in animals. "It's a complex task," Radosevic says. "You can't just chop off a part of a molecule and expect the rest to fold and function the same way."

IN THE POCKET

Recently, another target for a universal vaccine was discovered, one that is conserved but does not hide away in the protein folds. At the crown of hemagglutinin's globular head is the receptor-binding pocket, which allows the virus to infect host cells. With minor modifications, this pocket has the same structure on all the 16 subtypes of hemagglutinin. But the pocket is tiny — much smaller than a typical antibody. Nobody expected it to be a good vaccine target, and yet this appears to be the case.

To get snapshots of how the immune system reacts to the seasonal vaccine, researchers from Duke University in Durham, North Carolina, screened antibodies from volunteers several times after vaccination. One of these antibodies, dubbed CH65, seemed to be a little more effective than others. The Duke team sent the information to Stephen Harrison, a structural biologist at Harvard Medical School and Children's Hospital in Boston, Massachusetts, for analysis. "We were surprised to find that it docks to the receptor-binding pocket," says Harrison. "About half of the contacts are with amino acid residues in the pocket itself. The degree and intimacy of mimicry is striking."

What's more, point mutations in the area surrounding the pocket did not hinder the antibody from docking. When the researchers tested the antibody against 36 strains of H1 that had circulated within the last three decades, it blocked 30 of them. "Over 30 years of evolution of the virus, there were hardly any mutations that had a strong effect on binding," Harrison says. Another possibility, he adds, is that since 1977, the antibody arose so rarely that the virus did not have to escape its attack.

Harrison's research interest is not so much the creation of a universal flu vaccine as in advancing the basic science of affinity maturation: When exposed to the same antigens time and again, the immune system produces antibodies with slight mutations that will bind much more effectively to the invader. But why do some small changes make such a difference to the immune response — and how can one induce them?

"We boost what people already have. We are not trying to prime new responses."

While some scientists are busy mapping the conserved regions on the hemagglutinin molecule, others are still investigating existing vaccine targets that had for many years been the only ones available. These include the ion channel M2, the matrix protein 1 (M1) and the nucleoprotein that the virus needs to keep its genome stable. "We have known about these conserved proteins for at least 20 years," says Rappuoli. While acknowledging that "theoretically" one might prove to be a good target for a universal vaccine, he says that "there hasn't been much progress" on these fronts.

Many companies are still pursuing these targets. VaxInnate, a biotech company in Cranbury, New Jersey, has engineered a hybrid molecule that consists of four copies of M2e — part of the ion channel M2 that sits on the virus surface — fused into the bacterial protein flagellin. VaxInnate recently reported that its vaccine, called Vax102, safely produces an immune response in humans that should be protective against all strains of influenza A. DynaVax, a biotech company in Berkeley, California, has fused M2e and nucleoprotein to create a vaccine candidate called N895; the goal is to encourage antibodies against M2e as well as T cells against the nucleoprotein. Acambis, the UK biotech company based in Cambridge, now owned by Sanofi Pasteur of Lyon, France, evaluated its own M2e vaccine in a phase I study in 2008 and found that it was immunogenic against influenza A, and well-tolerated. The company also conducted another study that it says showed the vaccine protecting about 70% of ferrets against bird flu.

Further development has stagnated. Sanofi Pasteur is hesitant about moving its M2e vaccine to phase II trials. "M2e alone is unlikely to be better then the current seasonal vaccines," says Jeffrey Almond, Sanofi's vice president for discovery research and external R&D. "So we stepped back a little and thought about what we could add to M2e. There is no reason why a universal vaccine shouldn't have multiple components." Meanwhile, Merck, based in New Jersey, has halted its own trials.

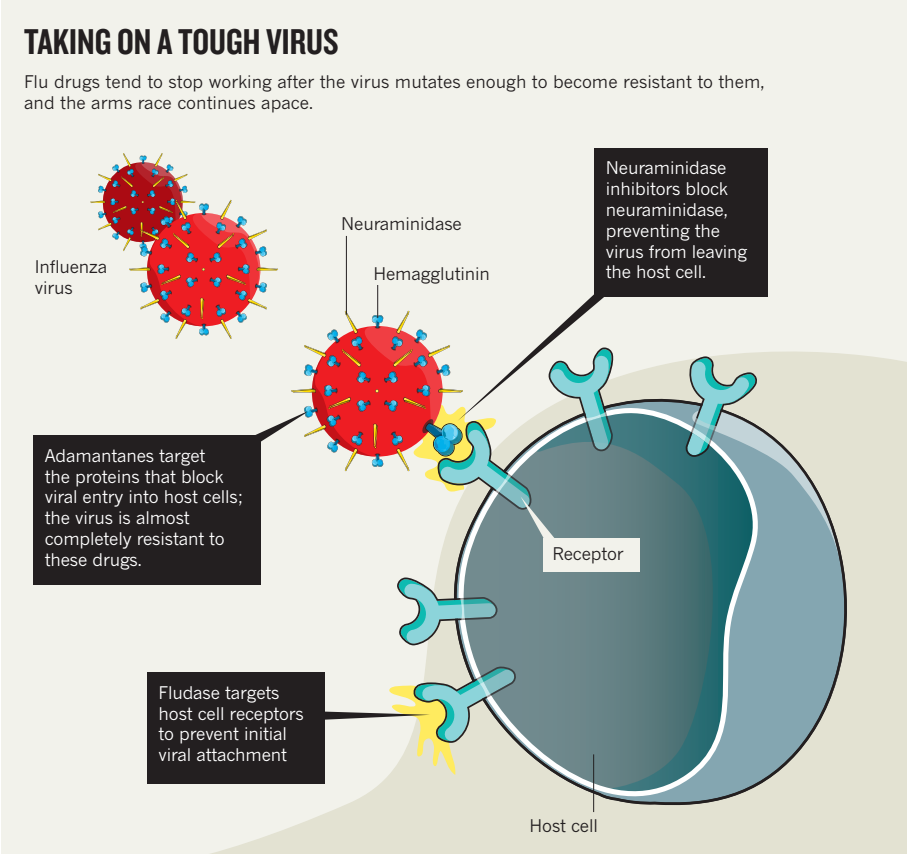
A DIFFERENT TACK

Another path toward a universal vaccine involves not trying to elicit antibodies, but rather boosting the body's T-cell response to infection. While antibodies prevent the virus from infecting host cells, T cells help to clear the virus from the body by killing flu-infected cells. That's the approach being pursued by vaccinologist Sarah Gilbert at the Jenner Institute — a vaccine research organization based at Oxford University in the UK. Gilbert uses an attenuated poxvirus called MVA, which presents the flu nucleoprotein and M1 to the immune system. As a result, Gilbert reports, "we saw a very large peak in T-cell response — everybody improved." This T-cell-mediated immunity might provide another line of defense against flu, possibly in combination with a protein vaccine that targets the stem of the hemagglutinin. Gilbert invited 11 vaccinated volunteers and 11 non-vaccinated volunteers to a quarantine facility. To test their vaccine, the researchers dripped H3N2 into the noses of their subjects and monitored their flu symptoms as well as their T-cell response.

Combining Gilbert's vaccine with the seasonal flu shot yields a preparation that boosts not only T cells but also the antibody response. Following vaccination with the seasonal flu shot alone, young people respond well by producing antibodies, whereas the same vaccine is less effective in older people. "Using both vaccines together might be extremely useful for vaccinating the elderly," Gilbert says. Usually, the seasonal flu jabs don't work as well for older people, because their immune response diminishes quickly, and it becomes more and more difficult to make new antibodies as we age. "We boost what people already have," Gilbert explains. "We are not trying to prime new responses."

Even a potent universal protein vaccine modelled after the perfect antibody would not change the fact that different age groups react differently to a flu jab. Thus a once-in-a-lifetime shot that protect against all strains might never be achieved for everyone: "The question is what universal means," says Rappuoli. "Should it cover all the pandemic strains and the seasonal strains? That's almost impossible. But it's OK to have a dream to move forward." ■

Jana Schlütter *is a science writer in Berlin.*



DRUGS

Lines of defence

Antiviral treatments are a critical component of an effective healthcare response to influenza, but drug resistance to the treatment-of-choice has public health officials searching for other options.

BY ROXANNE PALMER

The influenza virus has adapted to render the first generation of flu antivirals impotent, and there are signs of resistance to newer remedies. In late August 2011, the International Society for Infectious Diseases, based in Boston, Massachusetts, reported that 25 influenza patients in eastern Australia were infected with a strain of the notorious swine flu — influenza A (H1N1) — resistant to the widely used drug oseltamivir (Tamiflu), made by Roche, headquartered in Basel, Switzerland. The Australian case is the largest cohort of oseltamivir-resistant flu yet reported as scientists and pharmaceutical companies explore ways to outsmart the virus.

The good news is that resistance to antiviral drugs is not widespread in influenza strains for the upcoming 2011–2012 season — at least not yet. "Of the circulating strains of influenza in humans, we do not see resistance to oseltamivir," says Charles Penn, an antiviral drug expert with the global influenza program of the World Health Organization (WHO) in Geneva, Switzerland. Although resistance to oseltamivir was reported in as many as 1% of H1N1 flu samples collected during the winter of 2010–2011, says Penn, that level doesn't pose enough of a risk to warrant the WHO to change its treatment recommendations.

That doesn't mean that drug resistance is not a problem. Oseltamivir has successfully treated millions of patients since 1999, but mutations that confer resistance were described as early as 1998 — and given the ever-changing nature of flu, it can be hard to predict the trajectory of oseltamivir resistance. "In a span of two to three years we've seen a blossoming of resistance. At any time, the whole game can change," says Zachary Taylor, an infectious disease fellow at the Kaiser Permanente Fontana Medical Center in Sacramento, California. In part to safeguard against the possibility of such game-changing developments, drug developers are slowly filling the pipeline with alternative therapies (see 'Drugs to treat influenza infection'). Each drug come with side effects, which make them only worthwhile for those whom the flu could be potentially lethal — the elderly and the immunocompromised.

Given the wily history of the influenza virus, any sudden appearance of drug resistance is certain to concern public health officials. The first antiviral drugs to combat the disease — the adamantanes, which target the M2 channel protein to block virus entry into host cells — are now essentially useless. The US Centers for Disease Control and Prevention (CDC) found that 100 % of seasonal H3N2 flu in the 2009–2010 season and 99.8% of 2009 pandemic H1N1 flu were resistant to adamantanes.

Oseltamivir belongs to a class of drugs called neuraminidase inhibitors. These agents block the active site of a viral protein called neuraminidase (N), thereby arresting the influenza virus' ability to leave the host cell after it proliferates. The most common way for the influenza virus to evade oseltamivir is via the H275Y mutation (also known as H274Y) of neuraminidase, which replaces a single histidine amino acid with a tyrosine. This alteration interferes with the drug's ability to bind to the protein — a problem acknowledged by the maker of oseltamivir. "There remains a medical need and room for additional treatment options, especially for the management of severe infections and for improved pandemic preparedness," says Klaus Klumpp, Roche's top virologist. Klumpp says the Roche is supporting research into new therapies targeting viral replication as well as other mechanisms, but notes that these efforts are preclinical.

Fortunately, viruses with the H275Y mutation are still susceptible to a different neuraminidase inhibitor: zanamivir (marketed by UK-based GlaxoSmithKline (GSK) as Relenza). Zanamivir, the first neuraminidase inhibitor discovered, is generally administered by oral inhalation; its side effects include dizziness and nose irritation.

The WHO recommends using zanamivir to treat patients afflicted with oseltamivir-resistant flu strains, and it's already a viable alternative to oseltamivir. Zanamivir was actually the first neuraminidase inhibitor on the market, but its cousin oseltamivir was approved shortly afterwards and captured a larger market share. In January 2011, GSK began a double-blind study comparing intravenous zanamivir with oral oseltamivir. The study, which has enrolled 462 adolescents and adults, is due to be completed in September 2013.

While zanamivir remains an arrow in the quiver to be used in case of an oseltamivir-resistant strain, another drug

NATURE.COM
For some of the latest research into drugs for influenza:
go.nature.com/wkjvpp

DRUGS TO TREAT INFLUENZA INFECTION

Clinical name	Type of drug	Brand name and manufacturer	Status
Oseltamivir (oral)	Neuraminidase inhibitor	Tamiflu (Roche)	Commercially available in most countries
Zanamivir (inhaled)	Neuraminidase inhibitor	Relenza (GSK)	Commercially available in most countries
Peramivir (intravenous)	Neuraminidase inhibitor	Rapiacta, in Japan); Peramiflu in South Korea (BioCryst)	Available in Japan and South Korea; in Phase III testing in U.S.
DAS181 (inhaled)	Fusion protein	Fludase (NexBio)	Phase II testing in U.S.
ADS-8902 (adamantine, ribavarin, oseltamivir) (oral)	Triple combination	N/A (Adamas)	Phase II testing in U.S.
Amantadine (oral)	Adamantane	Symmetrel (Endo Pharmaceuticals)	Commercially available; not recommended for influenza due to resistance
Rimantidine (oral)	Adamantane	Flumadine (Forest Pharmaceuticals)	Commercially available; not recommended for influenza due to resistance

targeting viral neuraminidase is already available in Asia and might be useful for patients who can't tolerate the other neuraminidase inhibitors. This third neuraminidase inhibitor — peramivir, developed by BioCryst Pharmaceuticals of Durham, North Carolina — is available in Japan (as Rapiacta) and South Korea (as Peramiflu) but is still undergoing clinical trials in the United States. Its known side effects are similar to those of oseltamivir — diarrhea, nausea and vomiting. As an intravenous drug, peramivir can be administered to very sick or hospitalized patients who wouldn't be able to swallow oseltamivir or inhale zanamivir. In October 2009, the US Food and Drug Administration issued an emergency use authorization (EUA) for peramivir for people unable to take oseltamivir or zanamivir. The peramivir EUA expired in June 2010, and although the approval process has been fast-tracked in the United States, general approval of the drug is still several years away; the phase III clinical trial is not expected to be completed until May 2013.

THE VIRTUE OF VARIETY

Other pharmaceutical companies, cognizant of fears of resistance to neuraminidase inhibitors, are exploring a variety of ways to attack flu. One drug making its way along the pipeline is DAS 181, an oral antiviral drug developed by NexBio in San Diego, California, under the trade name Fludase, which is now in phase II clinical trials. The drug is a fusion protein, which is created by combining genetic sequences that encode two or more other proteins into a single protein. It aims to prevent infection by inactivating viral receptors on cells in a patient's respiratory tract, making it harder for the virus to latch onto host cells. A study by NexBio published in 2009 reports that Fludase was effective against 11 oseltamivir-resistant strains of the H1N1 virus — including a couple of variations that showed signs of reduced sensitivity to zanamivir. In September 2011, NexBio presented results from the phase II trial at the Interscience Conference on Antimicrobial Agents and Chemotherapy in Chicago, Illinois, that showed Fludase had cut the level of virus in patients' blood after just the second day of treatment, much as zanamivir and oseltamivir do. However, viral load has not been found to correlate with clinical severity of a flu patient's illness, and its use as a surrogate endpoint is controversial.

Any single flu drug has drawbacks. A mixture of treatments might be a better strategy to combat the virus — and in some cases, overcome resistance. Adamas Pharmaceuticals, based in Emeryville, California, for example, has been developing a three-drug combination treatment strategy of oseltamivir, amantadine — an adamantane drug thought to be obsolete — and ribavirin (another antiviral drug commonly used to treat hepatitis C infection). In cell cultures, the three drugs worked better than any pair, and were even able to inhibit viral activity in strains of influenza resistant to oseltamivir.

Nobody knows for sure how the three drugs work together. Adamas researcher Jack Nguyen speculates that each drug can take out the viruses that slipped past the others — a one-two-three punch combo. "After each stage you have fewer and fewer viruses making it through the cycle," Nguyen says. It may also be possible that oseltamivir's binding to neuraminidase causes other proteins on the virus' surface to change shape and become more susceptible to the other drugs, Nguyen adds.

In April 2010, Adamas released the results of a pilot study of its three-drug combination in seven patients with weakened immune systems. Five of the six patients that received the triple treatment responded by day 10 of therapy. The lone patient that was given only oseltamivir did not respond after 20 days. "This pilot study was an important first step in validating that the combination of three antivirals can provide a virologic and clinical benefit to patients at risk for complications of influenza," says Janet Englund, a clinical investigator who led the trial at Seattle Children's Hospital in Washington. The US National Institute of Allergy and Infectious Diseases in Bethesda, Maryland is recruiting 720 people for a phase II trial testing the efficacy of Adamas' triple combination therapy versus oseltamivir in patients with severe medical conditions, such as heart or lung disease, which make them more susceptible to serious complications from a bout of the flu.

More options are being explored. Recently researchers at the Scripps Research Institute in La Jolla, California, infected mice with a fatal dose of influenza, then gave some of the animals compounds that inhibited their ability to produce cytokines — cell signalling molecules that summon T cells and other immune system effectors to the site of infection. Mice that received these cytokine blockers plus the antiviral drug oseltamivir had a much higher survival rate (96%) than the other experimental groups that received only the cytokine-blocking compound (82% survival); oseltamivir alone (50%); or no treatment at all (21%). This suggests that much of the damage caused by flu is not due to the virus itself, but to an overenthusiastic response from a person's immune system. 'Cytokine storms', triggered by an overproduction of immune signaling molecules, can cause significant damage to tissue. Indeed, the phenomenon is thought to be responsible for a large proportion of deaths during the 1918 influenza pandemic.

Still, targeting the patient instead of the virus has its risks. "When you start tweaking the immune system, you have to wonder about whether you're under correcting or overcorrecting the immune response," explains Dean Blumberg, head of paediatric infectious diseases at the University of California, Davis. Overcorrect, and you exacerbate the problem of cytokine storms; under correct, and you run the risk of leaving the body defenceless against pathogens besides influenza.

"In a span of two to three years we've seen a blossoming of resistance. At any time, the whole game can change."

Blumberg says some patients shrug at the idea of using antiviral drugs at all, figuring that once contracted, the disease will simply run its course. In most cases, it will. In fact, neuraminidase inhibitors usually only shorten the duration of illness by about one day, and that's if the drug is taken within 48 hours of the first sign of flu symptoms. Yet Blumberg says that personally, "as a doctor and as a parent" he usually recommends a course of antivirals, to cut down on the length of illness.

With no signs of an avian or swine flu pandemic looming, the need for a wide spectrum of antiviral treatments might seem over cautious. But when warring against viruses, carefully laid preparations made in peacetime may serve well in the thick of the fight. ■

Roxanne Palmer *is a science writer in Brooklyn, New York.*

LETTER

doi:10.1038/nature10612

Fast core rotation in red-giant stars as revealed by gravity-dominated mixed modes

Paul G. Beck[1], Josefina Montalban[2], Thomas Kallinger[1,3], Joris De Ridder[1], Conny Aerts[1,4], Rafael A. García[5], Saskia Hekker[6,7], Marc-Antoine Dupret[2], Benoit Mosser[8], Patrick Eggenberger[9], Dennis Stello[10], Yvonne Elsworth[7], Søren Frandsen[11], Fabien Carrier[1], Michel Hillen[1], Michael Gruberbauer[12], Jørgen Christensen-Dalsgaard[11], Andrea Miglio[7], Marica Valentini[2], Timothy R. Bedding[10], Hans Kjeldsen[11], Forrest R. Girouard[13], Jennifer R. Hall[13] & Khadeejah A. Ibrahim[13]

When the core hydrogen is exhausted during stellar evolution, the central region of a star contracts and the outer envelope expands and cools, giving rise to a red giant. Convection takes place over much of the star's radius. Conservation of angular momentum requires that the cores of these stars rotate faster than their envelopes; indirect evidence supports this[1,2]. Information about the angular-momentum distribution is inaccessible to direct observations, but it can be extracted from the effect of rotation on oscillation modes that probe the stellar interior. Here we report an increasing rotation rate from the surface of the star to the stellar core in the interiors of red giants, obtained using the rotational frequency splitting of recently detected 'mixed modes'[3,4]. By comparison with theoretical stellar models, we conclude that the core must rotate at least ten times faster than the surface. This observational result confirms the theoretical prediction of a steep gradient in the rotation profile towards the deep stellar interior[1,5,6].

The asteroseismic approach to studying stellar interiors exploits information from oscillation modes of different radial order n and angular degree l, which propagate in cavities extending at different depths[7]. Stellar rotation lifts the degeneracy of non-radial modes, producing a multiplet of $(2l + 1)$ frequency peaks in the power spectrum for each mode. The frequency separation between two mode components of a multiplet is related to the angular velocity and to the properties of the mode in its propagation region. More information on the exploitation of rotational splitting of modes may be found in the Supplementary Information. An important new tool comes from mixed modes that were recently identified in red giants[3,4]. Stochastically excited solar-like oscillations in evolved G and K giant stars[8] have been well studied in terms of theory[9–12], and the main results are consistent with recent observations from space-based photometry[13,14]. Whereas pressure modes are completely trapped in the outer acoustic cavity, mixed modes also probe the central regions and carry additional information from the core region, which is probed by gravity modes. Mixed dipole modes ($l = 1$) appear in the Fourier power spectrum as dense clusters of modes around those that are best trapped in the acoustic cavity. These clusters, the components of which contain varying amounts of influence from pressure and gravity modes, are referred to as 'dipole forests'.

We present the Fourier spectra of the brightness variations of stars KIC 8366239 (Fig. 1a), KIC 5356201 (Supplementary Fig. 3a) and KIC 12008916 (Supplementary Fig. 5a), derived from observations with the Kepler spacecraft. The three spectra show split modes, the spherical degree of which we identify as $l = 1$. These detected multiplets cannot have been caused by finite mode lifetime effects from mode damping, because that would not lead to a consistent multiplet appearance over several orders such as that shown in Fig. 1. The spacings in period between the multiplet components (Supplementary Fig. 7) are too small to be attributable to consecutive unsplit mixed modes[4] and do not follow the characteristic frequency pattern of unsplit mixed modes[3]. Finally, the projected surface velocity, $v \sin i$, obtained from ground-based spectroscopy (Table 1), is consistent with the rotational velocity measured from the frequency splitting of the mixed mode that predominantly probes the outer layers. We are thus left with rotation as the only cause of the detected splittings.

The observed rotational splitting is not constant for consecutive dipole modes, even within a given dipole forest (Fig. 1b and Supplementary Figs 3b and 5b). The lowest splitting is generally present for the mode at the centre of the dipole forest, which is the mode with the largest amplitude in the outer layers. Splitting increases for modes with a larger gravity component, towards the wings of the dipole mode forest. For KIC 8366239, we find that the average splitting of modes in the wings of the dipole forests is 1.5 times larger than the mean splitting of the centre modes of the dipole forests.

We compared the observations (Fig. 1b) with theoretical predictions for a model representative of KIC 8366239, as defined in the Supplementary Information. The effect of rotation on the oscillation frequencies can be estimated in terms of a weight function, called a rotational kernel (K_{nl}). From the kernels, it is shown that at least 60% of the frequency splitting for the $l = 1$ mixed modes with a dominant gravity component is produced in the central region of the star (Fig. 2). This substantial core contribution to mixed modes enables us to investigate the rotational properties of the core region, which was hitherto not possible for the Sun, owing to a lack of observed modes that probe the core region (within a radius $r < 0.2\,R_{\odot}$; ref. 15). The solar rotational profile is known in great detail for only those regions probed by pressure-dominated modes[16–18]. In contrast to these modes in the wings of the dipole forest, only 30% of the splitting of the centre mode originates from the central region of the star, whereas the outer third by mass of the star contributes 50% of the frequency splitting. By comparing the rotational velocity derived from the splitting of such pressure-dominated modes with the projected surface velocity from spectroscopy, we find that the asteroseismic value is systematically larger. This offset cannot be explained by inclination of the rotation axis towards the observer alone (Supplementary Tables 1 and 2), but originates from the contribution of the fast-rotating core (Fig. 2). Furthermore, internal non-rigid rotation leads to a larger splitting for modes in the wings of the dipole forest than for centre modes.

[1]Instituut voor Sterrenkunde, Katholieke Universiteit Leuven, 3001 Leuven, Belgium. [2]Institut d'Astrophysique et de Géophysique de l'Université de Liège, 4000 Liège, Belgium. [3]Institut für Astronomie der Universität Wien, Türkenschanzstraße 17, 1180 Wien, Austria. [4]Afdeling Sterrenkunde, Institute for Mathematics Astrophysics and Particle Physics (IMAPP), Radboud University Nijmegen, 6500GL Nijmegen, The Netherlands. [5]Laboratoire Astrophysique, Instrumentation et Modélisation (AIM), CEA/DSM—CNRS—Université Paris Diderot; Institut de Recherche sur les lois Fondamentales de l'Univers/Service d'Astrophysique (IRFU/Sap), Centre de Saclay, 91191 Gif-sur-Yvette Cedex, France. [6]Astronomical Institute 'Anton Pannekoek', University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands. [7]School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. [8]Laboratoire d'études spatiales et d'instrumentation (LESIA), CNRS, Université Pierre et Marie Curie, Université Denis Diderot, Observatoire de Paris, 92195 Meudon Cedex, France. [9]Observatoire de Genève, Université de Genève, 51 Ch. des Maillettes, 1290 Sauverny, Switzerland. [10]Sydney Institute for Astronomy (SIfA), School of Physics, University of Sydney 2006, Australia. [11]Department of Physics and Astronomy, Aarhus University, DK-8000 Aarhus C, Denmark. [12]Department of Astronomy and Physics, Saint Marys University, Halifax, NS B3H 3C3, Canada. [13]Orbital Sciences Corporation/NASA Ames Research Center, Moffett Field, 94035 California, USA.

©2011 Macmillan Publishers Limited. All rights reserved

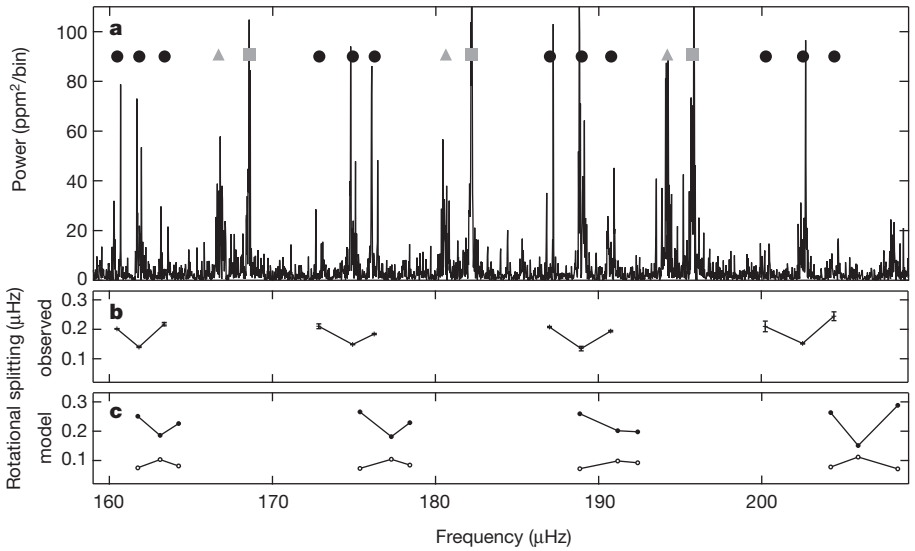


Figure 1 | Oscillation spectrum of KIC 8366239. a, Grey squares, radial modes; grey triangles, $l = 2$ modes; black circles, $l = 1$ rotational multiplets. A zoom on the region of 185–195 μHz and the analysis of the comb-like structure of the oscillation spectrum in an échelle diagram are shown in the online material (Supplementary Figs 1 and 2). The spectral window of the Fourier analysis can be found in the online material (Supplementary Fig. 11). The *y*-axis indicates the flux variation power in parts-per-million squared for each frequency bin. **b**, The observed rotational splitting for individual $l = 1$ modes. Error bars, standard deviation of the measured rotational splitting of dipole modes. Similar analyses of the stars KIC 5356201 (Supplementary Figs 3 and 4) and KIC 12008916 (Supplementary Figs 5 and 6) are discussed in the online material. **c**, Theoretically predicted rotational splitting assuming two different rotation laws. The values are calculated for a representative model of KIC 8366239 as defined in the Supplementary Information. Solid circles, splitting for non-rigid rotation for the case of a core rotation ten times faster than the surface rotation of 2.5 km s^{-1} resembles the observations qualitatively well. Open circles, theoretical splittings for rigid rotation and an equatorial surface rotation velocity of 3 km s^{-1} show a trend opposite to the observed one, with the largest splitting in the centre of the dipole forest and lower splitting in gravity-dominated modes. In the case of rigid rotation, the variable splitting is governed purely by the variation of the Ledoux[24] constant across the dipole forest (Supplementary Fig. 8). Because the representative model (**c**) has not been corrected for surface effects, there is a slight offset to the observations (**b**).

For a model rotating rigidly, the reverse behaviour is expected (Fig. 1c). Because the prediction for the rigidly rotating model is incompatible with the observed trends of the splittings (Fig. 1b) but can be well reproduced qualitatively under the assumption of non-rigid rotation (Fig. 1c), we conclude that the three stars investigated here (see Table 1, last column) rotate non-rigidly, with the central region rotating much faster than the surface.

The above interpretation is consistent with the correlation between the mode lifetime and the corresponding rotational splitting that has been observed in our data. Mixed modes in the wings of the dipole forest are predicted to have large amplitudes in the central regions of the star and, therefore, larger values of inertia and lifetime. These modes have narrower mode profiles in the frequency spectrum than have the centre modes that are predominantly trapped in the outer cavity[11]. This behaviour of the mode profiles (Supplementary Fig. 1)

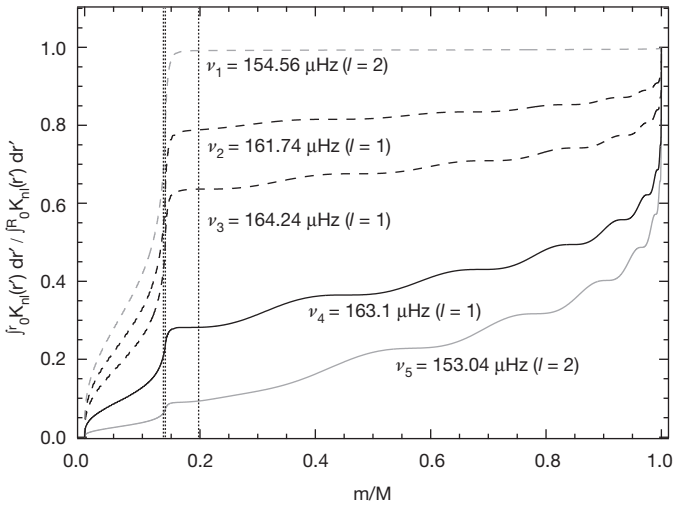


Figure 2 | Contributions to the total rotational splitting. Partial integrals of normalized rotation kernels, illustrating the contribution from different regions to the rotational splitting for pressure-dominated modes (ν_4, ν_5; solid lines) and gravity-dominated modes (ν_1, ν_2, ν_3; dashed lines) of degree $l = 1$ and $l = 2$, as a function of the stellar mass-fraction, m/M. The kernels have been calculated for modes from a representative model of KIC 8366239, as defined in the Supplementary Information, with oscillation frequencies given. Vertical dotted lines, left to right: boundary of the helium core, the hydrogen-burning shell and the bottom of the convective envelope.

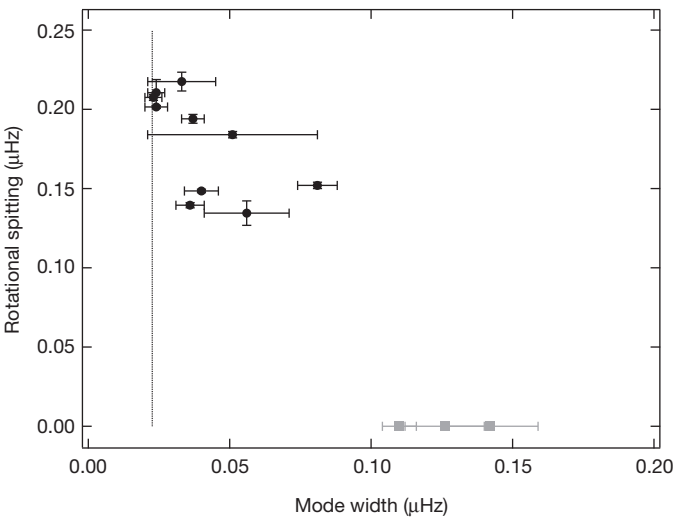


Figure 3 | Rotational splitting versus mode linewidth for KIC 8366239. The lifetime is inversely proportional to the mode linewidth. Black circles, $l = 1$ modes. Grey squares, linewidth of the pure acoustic radial modes ($l = 0$), for comparison. Dotted vertical line, formal frequency resolution. Error bars, standard deviation of the measured rotational splitting and mode width of dipole modes. Similar diagrams for the two other stars from Table 1 are shown in Supplementary Figs 9 and 10, respectively.

©2011 Macmillan Publishers Limited. All rights reserved

Table 1 | Observational parameters of three stars, showing rotational splitting.

KIC	ν_{max} (μHz)	R ($R_\odot$)	ΔP_{obs} (s)	T_{eff} (K)	$v \sin i$ (km s^{-1})	Asteros. rot. vel. (km s^{-1})	$l = 1$ min rot. split. (μHz)	Averaged split. ratio for dipole modes
8366239	182 ± 1	5.30 ± 0.08	56 ± 11	4980 ± 120	<1	3.9	0.135 ± 0.008	1.5
5356201	209.7 ± 0.7	4.47 ± 0.03	50 ± 10	4840 ± 90	2.4	3.8	0.154 ± 0.003	1.7
12008916	159.9 ± 0.6	5.18 ± 0.05	52 ± 7	4830 ± 100	-	5.7	0.200 ± 0.001	1.8

We present three cases of firm detections of rotational splitting in red giants from around 500 day-long time series of photometric data obtained with the Kepler satellite in long-cadence mode (about 30-min time sampling)[23]. KIC, stars' identifiers in the Kepler Input Catalog. ν_{max}, frequency of maximum oscillation power. The stellar radius R and the effective temperature T_{eff} were computed according to ref. 14. The mean period spacing ΔP_{obs} and its standard deviation were derived from the central multiplet components of the dipole modes. These three stars are classified as low-luminosity red-giant stars[22]. The observationally derived mean period spacing of the mixed modes indicates that they are in an early phase of red-giant evolution[4], in which they are burning hydrogen in a shell around the helium core. The projected surface velocity, $v \sin i$, was determined from ground-based spectra. The asteroseismic rotation velocity was computed from the minimum dipole splitting and assumed rigid rotation. Although this is the most pressure-dominated mode found in the spectrum, it contains a contribution from the core. The inferred rotation velocity from this mode is therefore slightly higher than the surface rotation velocity. The observed ratio of the average rotational splitting found in the wings of the dipole forest to the mean splitting of the forest centre is given in the last column. Extensive lists of further asteroseismic parameters are given in Supplementary Table 1 and 2.

was recently confirmed by observations[3,4]. We observe that short-lived modes (with broader profiles) exhibit smaller rotational splitting in KIC 8366239 (Fig. 3) and the two other stars from Table 1 (Supplementary Figs 9 and 10). With increasing lifetime (dominant gravity component) we see a substantial increase in the size of the rotational splitting of the modes. The frequencies of the narrowest modes are mainly affected by the rotation in the central regions of the star. Taking a representative model for KIC 8366239, and assuming that the convective envelope rotates rigidly and that the radiative interior rotates rigidly but faster than the convective envelope, we find that the observed splitting ratio of 1.5 (Table 1) is obtained from a core rotating at least ten times faster than the surface.

So far, our understanding of the evolution of the angular momentum as a function of the evolutionary stage is still poor. Before this study, it was possible to deliver evidence of non-rigid rotation in only two core-hydrogen-burning massive stars, for which the seismic analysis revealed a near-core rotation rate about three to five times faster than the envelope rotation rate[19,20]. Claims of non-rigid rotation in white dwarfs have also been discussed[21]. The low-luminosity red giants[22] discussed here are in the early phase of hydrogen-shell burning, and have a core-to-envelope rotation rate more than twice as high as those of the massive stars[19,20]. When these stars enter the subsequent phase of core-helium burning, we believe that the core will undergo a slight expansion and the envelope will shrink, leading to a surface rotation on the horizontal branch larger than that on the red-giant branch[1]. The basic assumption of conservation of angular momentum predicts a rotational gradient between surface and core that is less steep than the one detected. Exploiting rotational splitting of mixed modes in a large sample of red giants in various evolutionary stages will provide an excellent tool to inspect how the internal angular-momentum distribution evolves with time towards the end of the life of the star.

Received 29 July; accepted 29 September 2011.
Published online 7 December 2011.

1. Sills, A. & Pinsonneault, M. H. Rotation of horizontal-branch stars in globular clusters. *Astrophys. J.* **540,** 489–503 (2000).
2. Sweigart, A. V. & Mengel, J. G. Meridional circulation and CNO anomalies in red giant stars. *Astrophys. J.* **229,** 624–641 (1979).
3. Beck, P. G. *et al.* Kepler detected gravity-mode period spacings in a red giant star. *Science* **332,** 205 (2011).
4. Bedding, T. R. *et al.* Gravity modes as a way to distinguish between hydrogen- and helium-burning red giant stars. *Nature* **471,** 608–611 (2011).
5. Zahn, J.-P. Circulation and turbulence in rotating stars. *Astron. Astrophys.* **265,** 115–132 (1992).
6. Eggenberger, P. *et al.* Effects of rotation on the evolution and asteroseismic properties of red giants. *Astron. Astrophys.* **509,** A72 (2010).
7. Aerts, C., Christensen-Dalsgaard, J. & Kurtz, D. W. *Asteroseismology* Ch. 3 (Springer-Verlag, 2010).
8. De Ridder, J. *et al.* Non-radial oscillation modes with long lifetimes in giant stars. *Nature* **459,** 398–400 (2009).
9. Dziembowski, W. A. *et al.* Oscillations of α UMa and other red giants. *Mon. Not. R. Astron. Soc.* **328,** 601–610 (2001).
10. Christensen-Dalsgaard, J. Physics of solar-like oscillations. *Sol. Phys.* **220,** 137–168 (2004).
11. Dupret, M.-A. *et al.* Theoretical amplitudes and lifetimes of non-radial solar-like oscillations in red giants. *Astron. Astrophys.* **506,** 57–67 (2009).
12. Montalbán, J., Miglio, A., Noels, A., Scuflaire, R. & Ventura, P. Seismic diagnostics of red giants: first comparison with stellar models. *Astrophys. J.* **721,** L182–L188 (2010).
13. Mosser, B. *et al.* The universal red-giant oscillation pattern. An automated determination with CoRoT data. *Astron. Astrophys.* **525,** L9 (2011).
14. Kallinger, T. *et al.* Asteroseismology of red giants from the first four months of *Kepler* data: fundamental stellar parameters. *Astron. Astrophys.* **522,** A1 (2010).
15. García, R. A. *et al.* Tracking solar gravity modes: the dynamics of the solar core. *Science* **316,** 1591–1593 (2007).
16. Elsworth, Y. *et al.* Slow rotation of the Sun's interior. *Nature* **376,** 669–672 (1995).
17. Chaplin, W. J. *et al.* Rotation of the solar core from BiSON and LOWL frequency observations. *Mon. Not. R. Astron. Soc.* **308,** 405–414 (1999).
18. Thompson, M. J. *et al.* The internal rotation of the Sun. *Annu. Rev. Astron. Astrophys.* **41,** 599–643 (2003).
19. Aerts, C. *et al.* Asteroseismology of HD 129929: core overshooting and nonrigid rotation. *Science* **300,** 1926–1928 (2003).
20. Pamyatnykh, A. A. *et al.* Asteroseismology of the β Cephei star ν Eridani: interpretation and applications of the oscillation spectrum. *Mon. Not. R. Astron. Soc.* **350,** 1022–1028 (2004).
21. Kawaler, S. D., Sekii, T. & Gough, D. Prospects for measuring differential rotation in white dwarfs through asteroseismology. *Astrophys. J.* **516,** 349–365 (1999).
22. Bedding, T. R. *et al.* Solar-like oscillations in low-luminosity red giants: first results from *Kepler*. *Astrophys. J. Lett.* **713,** L176–L181 (2010).
23. García, R. A. *et al.* Preparation of *Kepler* light curves for asteroseismic analyses. *Mon. Not. R. Astron. Soc.* **414,** L6–L10 (2011).
24. Ledoux, P. The nonradial oscillations of gaseous stars and the problem of beta Canis Majoris. *Astrophys. J.* **114,** 373–384 (1951).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge the work of the team behind Kepler. Funding for the Kepler Mission is provided by NASA's Science Mission Directorate. P.G.B. and C.A. were supported by the European Community's Seventh Framework Programme (ERC grant PROSPERITY); J.D.R. and T.K. were supported by the Fund for Scientific Research, Flanders. S.H. was supported by the Netherlands Organisation for Scientific Research. J.M. and M.V. were supported by the Belgian Science Policy Office. The work is partly based on observations with the High Efficiency and Resolution Mercator Echelle Spectrograph at the Mercator Telescope, which is operated at La Palma in Spain by the Flemish Community.

Author Contributions P.G.B., T.K., J.D.R., C.A., R.A.G., S.H., B.M., Y.E., S.F., F.C. and M.G. measured the mode parameters, and derived and interpreted the rotational splitting and period spacings. J.M., M.-A.D., P.E., J.C.-D. and A.M. calculated stellar models and provided theoretical interpretation of the rotational splitting. M.H. and M.V. observed and analysed the spectra. J.D.R., S.H., S.F., Y.E., D.S., T.R.B., H.K., F.R.G., J.R.H. and K.A.I. contributed to the coordination of the project, including the acquisition and distribution of the data. C.A. defined and supervised the research. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.G.B. (paul.beck@ster.kuleuven.be).

©2011 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature10670

Open structure of the Ca^{2+} gating ring in the high-conductance Ca^{2+}-activated K^+ channel

Peng Yuan[1], Manuel D. Leonetti[1], Yichun Hsiung[1] & Roderick MacKinnon[1]

High-conductance voltage- and Ca^{2+}-activated K^+ channels function in many physiological processes that link cell membrane voltage and intracellular Ca^{2+} concentration, including neuronal electrical activity, skeletal and smooth muscle contraction, and hair cell tuning[1–8]. Like other voltage-dependent K^+ channels, Ca^{2+}-activated K^+ channels open when the cell membrane depolarizes, but in contrast to other voltage-dependent K^+ channels, they also open when intracellular Ca^{2+} concentrations rise. Channel opening by Ca^{2+} is made possible by a structure called the gating ring, which is located in the cytoplasm. Recent structural studies have defined the Ca^{2+}-free, closed, conformation of the gating ring, but the Ca^{2+}-bound, open, conformation is not yet known[9]. Here we present the Ca^{2+}-bound conformation of the gating ring. This structure shows how one layer of the gating ring, in response to the binding of Ca^{2+}, opens like the petals of a flower. The degree to which it opens explains how Ca^{2+} binding can open the transmembrane pore. These findings present a molecular basis for Ca^{2+} activation of K^+ channels and suggest new possibilities for targeting the gating ring to treat conditions such as asthma and hypertension.

Regulators of K^+ conductance (RCK) domains are ubiquitous among ion channels and transporters in prokaryotic cells[10–14]. Eight RCK domains assemble in the cytoplasm to form a closed-ring structure that changes its diameter on ligand binding, thus enabling the gating ring to regulate allosterically the transmembrane component of the transport protein. In prokaryotic cells, the gating ring most often consists of eight identical RCK domains, arranged as a tetrad of pairs, giving rise to a four-fold symmetric ring with identical top and bottom layers of RCK domains[12,15]. Intracellular ligands such as Ca^{2+} or small organic molecules bind in a cleft between RCK pairs forming the top and bottom layers to affect the shape of the ring[12,15–17].

RCK domains are also found in multicellular eukaryotes in the Slo family of K^+ channels[9,10,18,19]. There, two non-identical RCK domains are encoded in the carboxy terminus of the K^+ channel subunit. A tetrameric K^+ channel thus provides eight RCK domains to make a gating ring, but in contrast to most prokaryotic gating rings with identical RCK domains, the eukaryotic gating ring has one kind of RCK domain (RCK1) forming its top layer and another kind (RCK2) forming its bottom layer (Fig. 1a). The structure of a Ca^{2+}-activated (Slo1 or BK) K^+-channel gating ring in its Ca^{2+}-free, closed, conformation was recently determined (Protein Data Bank accession 3NAF), as was the structure of an RCK1–RCK2 pair in the presence of Ca^{2+} (PDB accession 3MT5), detailing the chemistry of the Ca^{2+}-binding site[9,18]. These studies showed that the BK gating ring indeed has distinct top and bottom layers, and also that a Ca^{2+}-binding site known as the Ca^{2+} bowl is in a different location than the ligand-binding site in prokaryotic gating rings. These unique structural properties suggest that to regulate conduction, the eukaryotic gating ring may undergo conformational changes in a different manner than its prokaryotic counterpart.

We determined the crystal structure of the Ca^{2+}-bound gating ring from the zebrafish BK channel at a resolution of 3.6 Å using an expression construct in which a loop consisting of residues 839–872 was deleted (Fig. 1b, Supplementary Table 1 and Supplementary Figs 1 and 2). This loop was disordered in both the Ca^{2+}-free human BK gating ring and the Ca^{2+}-bound RCK1–RCK2 pair (PDB accessions as above), and its deletion in both the human and zebrafish BK channels has no detectable affect on function[9,18] (Supplementary Fig. 3). Therefore, the relative stability of conformational states of the BK gating ring seem to be unaffected by this loop. Initial phases were determined by molecular replacement using the RCK1–RCK2 Ca^{2+}-bound structure as a search model[18]. The solution showed the presence of eight RCK1–RCK2 pairs (two gating rings) per asymmetric unit of the crystal, which allowed eight-fold non-crystallographic symmetry averaging of electron density maps and eight-fold restraints throughout model refinement. The first step of refinement required rigid-body adjustment of the amino-terminal lobe of RCK1. The final model was refined to working and free residuals (R_{work} and R_{free}) of 26.0% and 28.9%, respectively.

The zebrafish channel is 93% identical in amino-acid sequence to the human BK channel, for which the Ca^{2+}-free, presumably closed, conformation is known[9] (Fig. 1c and sequence alignment in Supplementary Information). The high degree of identity between zebrafish and human BK channels justifies comparison of the two structures to assess potential conformational changes underlying Ca^{2+}-mediated gating. The Ca^{2+}-bound and Ca^{2+}-free gating rings are shown in Fig. 1b, c. In these representations, RCK1, which forms the layer of the gating ring closest to the membrane, is coloured blue and RCK2 is coloured red. In both gating rings, a green asterisk marks the N-terminal residue of RCK1, Lys 343. In a full-length BK channel, the N-terminal residue of RCK1 is connected by a 16-amino-acid linker to the C terminus of the transmembrane inner helix, which forms the pore's gate[20].

The two gating ring structures, Ca^{2+}-bound and Ca^{2+}-free, differ in two main respects. First, in the Ca^{2+}-bound structure electron density attributable to Ca^{2+} is present in the Ca^{2+} bowl at the 'assembly interface', and, second, the RCK1 layer in the Ca^{2+}-bound gating ring is expanded from a diameter of 81 to 93 Å, measured at the position of Lys 343 (Fig. 1d, e and Supplementary Fig. 2c, d). An objective comparison of conformational differences between the Ca^{2+}-bound and Ca^{2+}-free gating rings is achieved by superimposing all Cα atoms of the tetrameric gating rings. Such a superposition shows that the RCK2 components of the ring undergo very little change, and that only modest conformational differences occur at the assembly interface, except locally near the Ca^{2+} bowl (Fig. 2c and Supplementary Fig. 4). The Ca^{2+} bowl structure is nearly identical to that of the Ca^{2+}-bound RCK1–RCK2 pair[18] (Supplementary Fig. 2e). The major conformational change involves the N-terminal lobe of RCK1, which changes its angle with respect to RCK2 (Fig. 2a, b and Supplementary Fig. 5). Viewing an interpolative movie in which the Ca^{2+}-free gating ring is morphed to the Ca^{2+}-bound gating ring, it seems that Ca^{2+} binding causes the N-terminal lobes of the RCK1 units to 'open up' on the membrane-facing surface of the gating ring in a way akin to petals opening on a flower (Supplementary Movie 1, part 1).

We next analysed whether these conformational changes are compatible with prior structural data on the opening and closing of K^+

[1]Laboratory of Molecular Neurobiology and Biophysics, Rockefeller University, Howard Hughes Medical Institute, 1230 York Avenue, New York, New York 10065, USA.

©2011 Macmillan Publishers Limited. All rights reserved

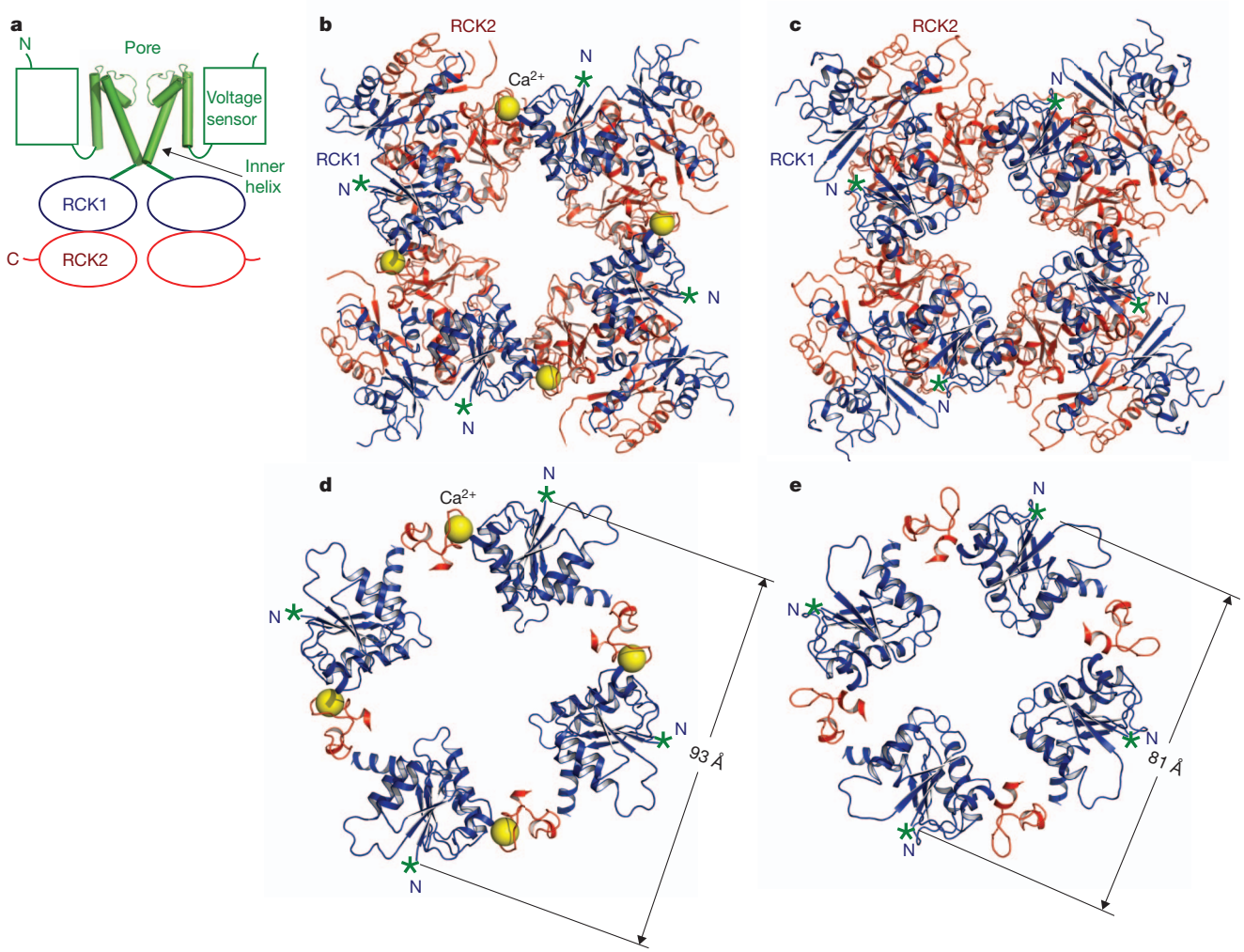


Figure 1 | The BK channel and the gating ring. **a**, Domain topology of the BK channel. For clarity, only two opposing subunits are shown. **b**, Crystal structure of the Ca^{2+}-bound, open, gating ring, with RCK1 in blue and RCK2 in red. The Ca^{2+} ions are shown as yellow spheres. The N termini of RCK1 that connect to the C termini of the inner helices are indicated as green asterisks. **c**, Structure of the Ca^{2+}-free, closed, gating ring, with RCK1 in blue and RCK2 in red (PDB accession 3NAF). **d**, The RCK1 N-terminal lobes (blue) and the Ca^{2+} bowls (red) from the Ca^{2+}-bound, open, gating ring. The diagonal distance between the Cα atoms of the N-terminal residues (Lys 343) is indicated. **e**, The corresponding region to that shown in **d** from the closed gating ring (PDB accession 3NAF).

channels. We started with the open prokaryotic K^+ channel MthK and replaced its gating ring with the Ca^{2+}-bound BK-channel gating ring[15]. This step involved alignment of the four-fold axis followed by minimization of the distance between the N terminus of the BK gating ring and the corresponding position within the MthK gating ring structure, through least-squares minimization of the distances between residues 343 of all four BK subunits and the corresponding residues, 116, of all four MthK subunits. The resulting model of an open BK-channel gating ring with an open K^+ pore is shown in Fig. 3a. The distance between the C-terminal residue of the inner helix of the pore and the N-terminal residue of RCK1, which was not constrained in making the model, is 33 Å. In a second step, we replaced the open MthK pore with the closed KcsA pore by aligning the invariant selectivity filter, and replaced the Ca^{2+}-bound BK-channel gating ring with the Ca^{2+}-free BK-channel gating ring by aligning the nearly invariant RCK2 layer[9,21]. The resulting model of a Ca^{2+}-free BK gating ring with a closed K^+ pore is shown in Fig. 3b. In this closed model, the distance between the C-terminal residue of the inner helix and the N-terminal residue of RCK1, which again was not constrained in making the model, is 32 Å. In this analysis, the similar pore-to-gating-ring distance (~32 Å) in the open model and the closed model suggests that the magnitude of conformational change observed within the gating ring is compatible with the magnitude of change known to occur in the pores of other kinds of K^+ channel (Supplementary Movie 1, part 2). Whether or not the pore of the BK channel undergoes conformational changes similar to or different than those of other K^+ channels remains to be determined[22–26].

Another feature of the model in Fig. 3 is noteworthy. In the published structure of the Ca^{2+}-free gating ring, part of the linker extending centrally from the gating ring towards the pore was present in the crystal structure and visible on the surface of the gating ring[9]. The linker was held by a four-helix bundle fused to the N terminus to facilitate crystallization. Although the linker position in the structure might be affected by this attachment, we note that it is almost coincident with the dashed lines in our closed model (Fig. 3b). The agreement between the observed linker position in the crystal structure of the closed gating ring and its inferred position in the closed conformation of our model is an interesting correlation because the radial direction of the linkers in the model (dashed lines) would naturally allow the expansion of the gating ring to exert an opening force on the pore's inner helices. The idea that the gating ring exerts a direct force on the pore has been proposed on the basis of mutagenesis studies showing that the linker length is critical to efficient Ca^{2+}-mediated gating[19,20].

The structural differences between the BK gating ring and the MthK gating ring underlie distinct ligand-induced conformational changes. A higher degree of molecular symmetry in the MthK gating ring stems from its containing eight identical RCK domains[12,15]. Consequently, the MthK gating ring changes equally on both layers such that Ca^{2+}

©2011 Macmillan Publishers Limited. All rights reserved

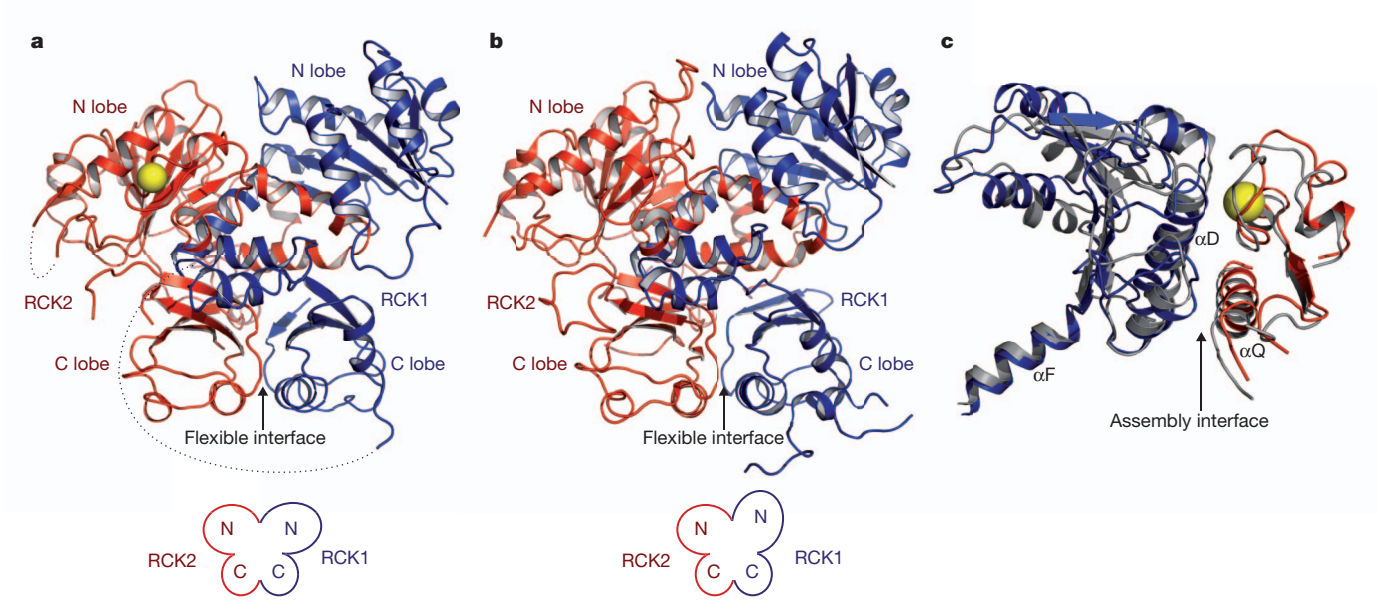


Figure 2 | The flexible interface and the assembly interface. **a**, The flexible interface from the open gating ring, with RCK1 in blue and RCK2 in red. The Ca^{2+} ions are shown as yellow spheres. The N- and C-terminal lobes of RCK1 and RCK2 are labelled. Large disordered regions are indicated by dotted lines. **b**, The flexible interface from the closed gating ring, with RCK1 in blue and RCK2 in red (PDB accession 3NAF). **c**, Superposition of the assembly interfaces from the open and closed gating ring structures. Secondary structural elements αF, αD and αQ are labelled.

binding causes a shrinking of the ring's height and an expansion of its diameter[12] (Supplementary Movie 2). By contrast, in the BK gating ring Ca^{2+} binding causes only the layer facing the membrane to undergo a substantial conformational change (Supplementary Movie 1). This comparison presents a fascinating example in which the evolution of molecular structure has given rise to new or modified mechanical properties within a class of molecules. Precisely how the free energy of Ca^{2+} binding in either the BK or MthK gating rings is transduced into

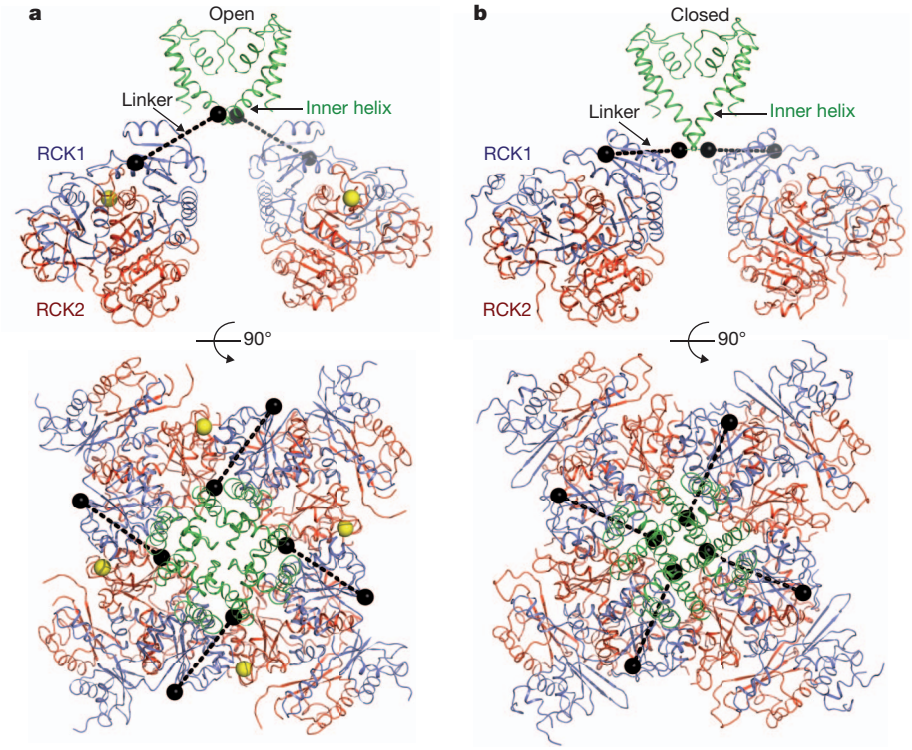


Figure 3 | A Ca^{2+}-gating model for the BK channel. **a**, Views of the BK-channel model with an open gating ring and an open pore. The structure of the Ca^{2+}-bound BK gating ring is docked onto the open MthK channel (PDB accession 1LNQ) by aligning the gating rings. Yellow spheres represent Ca^{2+} ions. The N-terminal residues of RCK1 and the C-terminal residues of the inner helices are shown as black spheres. **b**, Views of the BK channel model with a Ca^{2+}-free gating ring and a closed pore. Here, the Ca^{2+}-bound gating ring in **a** is replaced by the Ca^{2+}-free gating ring (PDB accession 3NAF) and the open MthK pore in **a** is replaced by the closed KcsA pore (PDB accession 1K4C). The inner helix from KcsA is truncated to match the length in MthK. See Supplementary Movie 1.

©2011 Macmillan Publishers Limited. All rights reserved

mechanical work remains an important outstanding question that will require further experimental and theoretical work to understand at a deep level.

The data and analysis presented here define end points in the Ca^{2+}-induced gating conformational transition of the BK gating ring. Given the large conformational change that occurs on Ca^{2+} binding, it should in principle now be possible to identify small molecules that stabilize one conformation or the other. Such small molecules promise to have important and possibly beneficial physiological effects because BK channels regulate smooth muscle tone in pulmonary airways and vascular beds[27–30].

METHODS SUMMARY

The cytoplasmic domain construct of the zebrafish BK channel including residues 341–1,060 with a loop deletion (residues 839–872) was expressed and purified as described previously[18]. However, the size exclusion chromatography was carried out in buffer containing 20 mM HEPES (pH 8.0), 150 mM NaCl and 20 mM dithiothreitol. The protein was concentrated to about 6 mg ml^{-1} and 10 mM $CaCl_2$ was added before crystallization experiments were performed. Crystals were grown at 20 °C using hanging-drop vapour diffusion by mixing equal volumes of protein and a reservoir solution containing 50 mM 2-(*N*-morpholino)ethanesulfonic acid, 4% (w/v) PEG 4000 and 100 mM potassium sodium tartrate (pH 6.3). The crystals belong to space group $P2_12_12_1$ and the unit cell has dimensions of $a = 137.65$ Å, $b = 210.82$ Å and $c = 238.76$ Å with $\alpha = \beta = \gamma = 90°$. Each asymmetric unit contains eight protein subunits forming two gating rings. The structure was determined by molecular replacement using the monomeric Ca^{2+}-bound human BK cytoplasmic domain (PDB accession 3MT5) as the search model. The final model was refined to a resolution of 3.6 Å with $R_{work} = 0.260$ and $R_{free} = 0.289$. Electrophysiology recordings were conducted using patch–clamp on *Xenopus oocytes* expressing the wild-type and mutant channels.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 July; accepted 24 October 2011.
Published online 4 December 2011.

1. Robitaille, R., Garcia, M. L., Kaczorowski, G. J. & Charlton, M. P. Functional colocalization of calcium and calcium-gated potassium channels in control of transmitter release. *Neuron* **11,** 645–655 (1993).
2. Fettiplace, R. & Fuchs, P. A. Mechanisms of hair cell tuning. *Annu. Rev. Physiol.* **61,** 809–834 (1999).
3. Nelson, M. T. *et al.* Relaxation of arterial smooth muscle by calcium sparks. *Science* **270,** 633–637 (1995).
4. Brenner, R. *et al.* Vasoregulation by the β1 subunit of the calcium-activated potassium channel. *Nature* **407,** 870–876 (2000).
5. Petkov, G. V. *et al.* β1-subunit of the Ca^{2+}-activated K^+ channel regulates contractile activity of mouse urinary bladder smooth muscle. *J. Physiol. (Lond.)* **537,** 443–452 (2001).
6. Lee, U. S. & Cui, J. BK channel activation: structural and functional insights. *Trends Neurosci.* **33,** 415–423 (2010).
7. Kaczorowski, G. J., Knaus, H. G., Leonard, R. J., McManus, O. B. & Garcia, M. L. High-conductance calcium-activated potassium channels; structure, pharmacology, and function. *J. Bioenerg. Biomembr.* **28,** 255–267 (1996).
8. Salkoff, L., Butler, A., Ferreira, G., Santi, C. & Wei, A. High-conductance potassium channels of the SLO family. *Nature Rev. Neurosci.* **7,** 921–931 (2006).
9. Wu, Y., Yang, Y., Ye, S. & Jiang, Y. Structure of the gating ring from the human large-conductance Ca^{2+}-gated K^+ channel. *Nature* **466,** 393–397 (2010).
10. Jiang, Y., Pico, A., Cadene, M., Chait, B. T. & MacKinnon, R. Structure of the RCK domain from the *E. coli* K^+ channel and demonstration of its presence in the human BK channel. *Neuron* **29,** 593–601 (2001).
11. Albright, R. A., Ibar, J. L., Kim, C. U., Gruner, S. M. & Morais-Cabral, J. H. The RCK domain of the KtrAB K^+ transporter: multiple conformations of an octameric ring. *Cell* **126,** 1147–1159 (2006).
12. Ye, S., Li, Y., Chen, L. & Jiang, Y. Crystal structures of a ligand-free MthK gating ring: insights into the ligand gating mechanism of K^+ channels. *Cell* **126,** 1161–1173 (2006).
13. Bakker, E. P., Booth, I. R., Dinnbier, U., Epstein, W. & Gajewska, A. Evidence for multiple K^+ export systems in *Escherichia coli*. *J. Bacteriol.* **169,** 3743–3749 (1987).
14. Roosild, T. P., Miller, S., Booth, I. R. & Choe, S. A mechanism of regulating transmembrane potassium flux through a ligand-mediated conformational switch. *Cell* **109,** 781–791 (2002).
15. Jiang, Y. *et al.* Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* **417,** 515–522 (2002).
16. Bellamacina, C. R. The nicotinamide dinucleotide binding motif: a comparison of nucleotide binding proteins. *FASEB J.* **10,** 1257–1269 (1996).
17. Dong, J., Shi, N., Berke, I., Chen, L. & Jiang, Y. Structures of the MthK RCK domain and the effect of Ca^{2+} on gating ring stability. *J. Biol. Chem.* **280,** 41716–41724 (2005).
18. Yuan, P., Leonetti, M. D., Pico, A. R., Hsiung, Y. & MacKinnon, R. Structure of the human BK channel Ca^{2+}-activation apparatus at 3.0 Å resolution. *Science* **329,** 182–186 (2010).
19. Pico, A. R. *RCK Domain Model of Calcium Activation in BK Channels* 43–46. PhD thesis, Rockefeller Univ. (2003).
20. Niu, X., Qian, X. & Magleby, K. L. Linker-gating ring complex as passive spring and Ca^{2+}-dependent machine for a voltage- and Ca^{2+}-activated potassium channel. *Neuron* **42,** 745–756 (2004).
21. Zhou, Y., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. Chemistry of ion coordination and hydration revealed by a K^+ channel–Fab complex at 2.0 Å resolution. *Nature* **414,** 43–48 (2001).
22. Wilkens, C. M. & Aldrich, R. W. State-independent block of BK channels by an intracellular quaternary ammonium. *J. Gen. Physiol.* **128,** 347–364 (2006).
23. Li, W. & Aldrich, R. W. State-dependent block of BK channels by synthesized shaker ball peptides. *J. Gen. Physiol.* **128,** 423–441 (2006).
24. Zhou, Y., Xia, X. M. & Lingle, C. J. Cysteine scanning and modification reveal major differences between BK channels and Kv channels in the inner pore region. *Proc. Natl Acad. Sci. USA* **108,** 12161–12166 (2011).
25. Tang, Q. Y., Zeng, X. H. & Lingle, C. J. Closed-channel block of BK potassium channels by bbTBA requires partial activation. *J. Gen. Physiol.* **134,** 409–436 (2009).
26. Li, W. & Aldrich, R. W. Unique inner pore properties of BK channels revealed by quaternary ammonium block. *J. Gen. Physiol.* **124,** 43–57 (2004).
27. Wu, R. S. & Marx, S. O. The BK potassium channel in the vascular smooth muscle and kidney: α- and β-subunits. *Kidney Int.* **78,** 963–974 (2010).
28. Ghatta, S., Nimmagadda, D., Xu, X. & O'Rourke, S. T. Large-conductance, calcium-activated potassium channels: structural and functional implications. *Pharmacol. Ther.* **110,** 103–116 (2006).
29. Miura, M., Belvisi, M. G., Stretton, C. D., Yacoub, M. H. & Barnes, P. J. Role of potassium channels in bronchodilator responses in human airways. *Am. Rev. Respir. Dis.* **146,** 132–136 (1992).
30. Jones, T. R., Charette, L., Garcia, M. L. & Kaczorowski, G. J. Interaction of iberiotoxin with β-adrenoceptor agonists and sodium nitroprusside on guinea pig trachea. *J. Appl. Physiol.* **74,** 1879–1884 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank staff members at NSLS X29, Brookhaven National Laboratory, for beamline assistance, and members of the MacKinnon laboratory for discussion. We thank P. Hoff and members of the Gadsby laboratory for help with oocyte preparation. R.M. is an investigator in the Howard Hughes Medical Institute. The research is supported by the American Asthma Foundation grant 07-0127.

Author Contributions P.Y. purified and crystallized the protein, collected the X-ray diffraction data, determined the structure and conducted electrophysiology recordings. M.D.L. aided in initial crystallization and electrophysiology experiments. Y.H. provided assistance with protein expression. P.Y. and R.M. designed the research and analysed data. P.Y., M.D.L. and R.M. prepared the manuscript.

Author Information Atomic coordinates and structure factors for the reported crystal structure have been deposited into the Protein Data Bank under accession code 3U6N. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.M. (mackinn@rockefeller.edu).

©2011 Macmillan Publishers Limited. All rights reserved

METHODS

Cloning, expression and purification. A synthetic gene encoding the zebrafish BK channel (GI: 189526846) was purchased from Bio Basic Inc. and served as the template for subcloning. The amino-acid sequence is 93% identical to that of the human BK channel. On the basis of previous structural and functional studies of the human BK channel[18], a cytoplasmic domain (CTD) construct including residues 341–1,060 with a loop deletion (residues 839–872) was expressed and purified following a method previously described[18]. However, the size exclusion chromatography was carried out in 20 mM HEPES (pH 8.0), 150 mM NaCl and 20 mM dithiothreitol. The protein was concentrated to about 6 mg ml^{-1} and 10 mM $CaCl_2$ was added to achieve the ligand-bound, open, conformation before crystallization experiments.

Crystallization and structure determination. Crystals of the zebrafish BK CTD were grown at 20 °C using hanging-drop vapour diffusion by mixing equal volumes of protein and a reservoir solution containing 50 mM 2-(*N*-morpholino)ethanesulfonic acid, 4% (w/v) PEG 4000 and 100 mM potassium sodium tartrate (pH 6.3). The crystals grew to a maximum size of about 0.2 mm × 0.2 mm × 0.3 mm within three days. Before being flash frozen in liquid nitrogen, the crystals were briefly transferred to a cryoprotectant solution containing 50 mM 2-(*N*-morpholino)ethanesulfonic acid, 6% (w/v) PEG 4000, 100 mM potassium sodium tartrate, 150 mM NaCl, 10 mM $CaCl_2$ and 30% ethylene glycol (pH 6.3). The crystals belong to space group $P2_12_12_1$ and the unit cell has dimensions of $a = 137.65$ Å, $b = 210.82$ Å and $c = 238.76$ Å with $\alpha = \beta = \gamma = 90°$. Each asymmetric unit contains eight protein subunits.

Diffraction data were measured at beamline X29 of the National Synchrotron Light Source and were processed with the HKL2000 program suite[31]. The scaled data set was anisotropically corrected to resolution limits of 4.0, 3.6 and 3.6 Å along the reciprocal cell directions a^*, b^*, and c^*, respectively, using the diffraction anisotropy server at the University of California, Los Angeles[32]. An isotropic B factor of −29.98 Å^2 was applied to restore the magnitude of the high-resolution reflections diminished by anisotropic scaling. The structure was determined by molecular replacement using the monomeric Ca^{2+}-bound human BK CTD (PDB accession 3MT5) as the search model[18]. Initial phases were improved by eight-fold non-crystallographic symmetry averaging carried out with DM from the CCP4 program suite[33]. Iterative model building was carried out in COOT[34]. Rounds of refinement were performed with CNS[35] and REFMAC[36], and strong eight-fold non-crystallographic symmetry restraints were maintained throughout refinement. The final model was refined to a resolution of 3.6 Å with $R_{work} = 0.260$ and $R_{free} = 0.289$. A few disordered regions were not modelled owing to weak electron density, and the final refined model includes residues 343–571, 581–614, 688–810, 817–837, 875–949, 953–1,024 and 1,029–1,059. Residues for which side-chain density was poorly defined were modelled as alanines. The majority (95.9%) of the residues lie in the most favoured region in a Ramachandran plot, with the remaining 4.1% in the additionally allowed regions. Eight Ca^{2+} ions were modelled in the Ca^{2+} bowls from the two gating rings in the crystal. Data collection and structure refinement statistics are shown in Supplementary Table 1. All structural illustrations were prepared with PYMOL (http://www.pymol.org). The multiple-chain morphing script for CNS[37] was used to generate coordinates for the movies in Supplementary Information.

Electrophysiology. Channel expression in *Xenopus laevis* oocytes and electrophysiological recordings in inside-out patches were carried out as previously described[18]. Wild-type and deletion zebrafish BK constructs were cloned into the pGEM-HE vector[38]. Plasmids were linearized with PciI and capped cRNA was produced by *in vitro* transcription from the T7 promoter using the AmpliCap-Max T7 kit (Epicentre Biotechnologies). *Xenopus laevis* oocytes were injected with ~10 ng of cRNA and recordings were made 48–60 h later. For recording, inside-out patches were excised from freshly devitellinized oocytes using fire-polished glass pipettes with a typical resistance of 1–2 MΩ. In Ca^{2+}-free conditions, the bath solution contained 140 mM K-$MeSO_3$, 20 mM HEPES, 2 mM KCl, 5 mM EGTA (final pH, 7.0). For recordings in 10 μM free Ca^{2+}, the same solution was supplemented with 4.83 mM $CaCl_2$, adjusting the pH to 7.0 (we assumed a K_d of 376 nM for the EGTA–Ca^{2+} complex at pH 7.0). The pipette solution was the same as the Ca^{2+}-free bath solution, but supplemented with 2 mM $MgCl_2$. During recording, the membrane voltage was held at −60 mV and currents were elicited by pulsing the voltage from −120 to +140 mV in 10-mV increments (length of pulse, 20 ms). For plots of conductance versus voltage, measurements were taken from tail currents at –60 mV, 700 μs after the voltage step.

31. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
32. Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103,** 8060–8065 (2006).
33. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D* **59,** 1131–1137 (2003).
34. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
35. Brunger, A. T. Version 1.2 of the crystallography and NMR system. *Nature Protocols* **2,** 2728–2733 (2007).
36. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53,** 240–255 (1997).
37. Echols, N., Milburn, D. & Gerstein, M. MolMovDB: analysis and visualization of conformational change and structural flexibility. *Nucleic Acids Res.* **31,** 478–482 (2003).
38. Liman, E. R., Tytgat, J. & Hess, P. Subunit stoichiometry of a mammalian K^+ channel determined by construction of multimeric cDNAs. *Neuron* **9,** 861–871 (1992).

©2011 Macmillan Publishers Limited. All rights reserved

ARTICLE

doi:10.1038/nature10674

A disinhibitory microcircuit for associative fear learning in the auditory cortex

Johannes J. Letzkus[1]*, Steffen B. E. Wolff[1,2]*, Elisabeth M. M. Meyer[1,2], Philip Tovote[1], Julien Courtin[3], Cyril Herry[3] & Andreas Lüthi[1]

Learning causes a change in how information is processed by neuronal circuits. Whereas synaptic plasticity, an important cellular mechanism, has been studied in great detail, we know much less about how learning is implemented at the level of neuronal circuits and, in particular, how interactions between distinct types of neurons within local networks contribute to the process of learning. Here we show that acquisition of associative fear memories depends on the recruitment of a disinhibitory microcircuit in the mouse auditory cortex. Fear-conditioning-associated disinhibition in auditory cortex is driven by foot-shock-mediated cholinergic activation of layer 1 interneurons, in turn generating inhibition of layer 2/3 parvalbumin-positive interneurons. Importantly, pharmacological or optogenetic block of pyramidal neuron disinhibition abolishes fear learning. Together, these data demonstrate that stimulus convergence in the auditory cortex is necessary for associative fear learning to complex tones, define the circuit elements mediating this convergence and suggest that layer-1-mediated disinhibition is an important mechanism underlying learning and information processing in neocortical circuits.

The transformation of sensory input to an appropriate behavioural output is accomplished by neuronal circuits that process information through dynamic interactions between distinct types of neurons. Learning as an adaptive change of an animal's behaviour modifies these network computations. Important cellular mechanisms underlying learning are thought to be activity-dependent synaptic plasticity[1] and subsequent structural modifications[2]. In contrast to our detailed understanding of these cellular mechanisms, the factors and circuit elements controlling neuronal activity and concomitant induction of plasticity during learning in the intact animal remain poorly understood.

Neocortical interneurons exert powerful control over circuit activity by supplying inhibition to surrounding pyramidal neurons and to other interneurons[3]. The balance of excitation and inhibition is critical to circuit function, and is maintained under most conditions including sensory stimulation, when thalamocortical input typically recruits feed-forward inhibition a few milliseconds after direct excitation[4,5]. This form of inhibition is mainly supplied by basket cells, the most abundant type of interneuron in the rodent neocortex[3]. Fast-spiking basket cells establish strong synapses with high release probability on the perisomatic region of pyramidal neurons where they control firing, and express the calcium-binding protein parvalbumin (PV)[3,6,7].

On top of the glutamatergic recruitment of inhibition by sensory input, several types of interneurons are also major targets of neuromodulation[8–11], suggesting that these systems can profoundly affect circuit computations by shifting the excitation–inhibition balance[12]. This has mainly been addressed *in vitro*, revealing a staggering complexity of effects which strongly depend on interneuron type[8–11]. Given that neuromodulation has a key role in learning[13,14], it is likely that interneurons are engaged in these processes. However, the contribution of different interneuron types to learning in the intact animal remains elusive.

Auditory fear conditioning is a form of associative learning acquired by temporal coincidence of a neutral conditioned stimulus (CS, a tone) with a mild foot shock. Fear conditioning causes prominent, long-lasting plasticity of CS responses in auditory cortex[13–15], which depends on the activity of cholinergic afferents from the basal forebrain[13,14,16]. Pairing of tones with basal forebrain stimulation can elicit similar plasticity in auditory cortex[13,14], and a recent study suggests an involvement of pyramidal neuron disinhibition in this process[17]. Here, we investigate the contribution of identified types of auditory cortex interneurons to acquisition of auditory fear memory. Our results indicate that foot shocks, via recruitment of cholinergic basal forebrain afferents, activate a disinhibitory neocortical microcircuit that is required for fear learning.

Nicotinic activation of L1 interneurons by foot shocks

Fear memory acquisition critically depends on the amygdala[18]. In contrast, the role of auditory cortex is contentious, with evidence both for[19,20] and against an essential function in auditory fear conditioning[19,21]. Based on the hypothesis that stimulus complexity determines the engagement of neocortex[18], we performed differential fear conditioning using trains of upward and downward frequency-modulated sweeps as the CS (Fig. 1a, modulation between 5 and 15 kHz). To test directly the role of auditory cortex in this paradigm, we inhibited neuronal activity during fear conditioning by local injection of the $GABA_A$ (γ-aminobutyric acid)-receptor agonist muscimol into the auditory cortex (Fig. 1a and Supplementary Fig. 1). When tested 24 h later in a drug-free state, this manipulation resulted in strongly reduced fear levels (Fig. 1b), indicating that activity in auditory cortex is required for fear learning in this paradigm.

We next focused on the circuit mechanisms underlying acquisition of fear memory in the auditory cortex (defined here as areas A1 and AuV). The role of the auditory cortex could be that of an essential relay for tone information to downstream structures like the amygdala. Alternatively, foot shocks alone might also elicit activity in the auditory cortex, and convergence of tone and foot-shock information during fear conditioning could then contribute to fear learning. To distinguish these alternatives, we used two-photon calcium imaging of neurons in

[1]Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland. [2]University of Basel, CH-4003 Basel, Switzerland. [3]INSERM U862, Neurocentre Magendie, 146 Rue Léo-Saignat, 33077 Bordeaux, France.
*These authors contributed equally to this work.

©2011 Macmillan Publishers Limited. All rights reserved

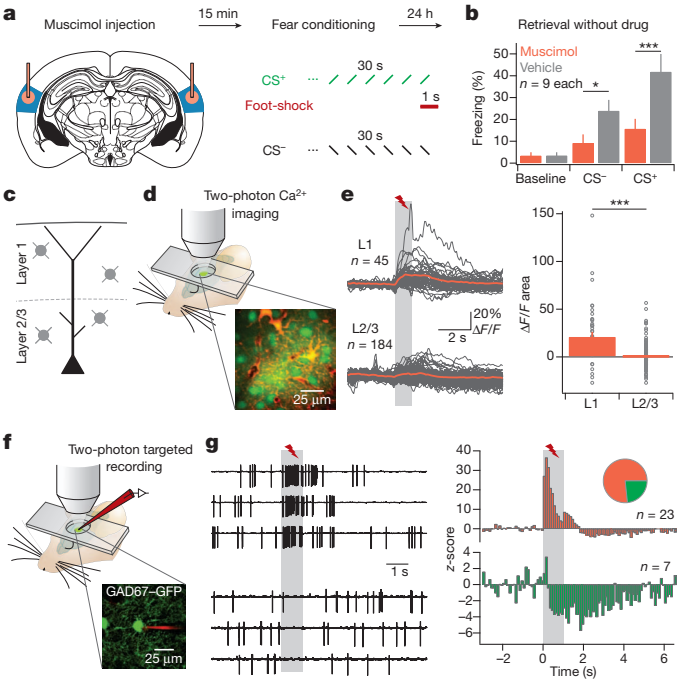


Figure 1 | Foot-shock responses in auditory cortex L1 interneurons. **a**, Left, injection of the $GABA_A$-receptor agonist muscimol (red) into the auditory cortex (blue). Right, differential fear-conditioning protocol using frequency-modulated sweeps. **b**, Drug-free freezing 1 day after conditioning in a new context. Compared to vehicle-injected mice (grey), muscimol reduced freezing to both CS^- and CS^+ (red). **c**, Cytoarchitecture of upper layers of auditory cortex (interneurons grey, pyramidal neuron black). **d**, Two-photon calcium imaging in head-fixed, anaesthetized mice using OGB-1 AM (green) and sulforhodamine 101 (red, counterstains glial cells). **e**, Left, responses in L1 and L2/3 to foot shocks in single neurons (grey) and the population average (red). Right, L1 interneurons display much stronger activation than L2/3 neurons. **f**, Two-photon targeted loose-seal cell-attached recording of L1 interneuron (green). **g**, Example traces (left) and *z*-scored population peri-stimulus time histograms (right) of two types of foot-shock responses in L1 interneurons. Inset, incidence of response type. Values are mean ± s.e.m. $*P < 0.05$, $***P < 0.001$. Statistical analysis in Supplementary Information.

the superficial layers of the auditory cortex in mice anaesthetized with chlorprothixene and urethane (Fig. 1c, d). Trains of frequency-modulated sweeps used as the CS strongly activated neurons in layer 2/3 (L2/3, Supplementary Fig. 2), whereas mild foot shocks applied to the hindpaws evoked little response (Fig. 1e and Supplementary Fig. 2). In marked contrast, the population of neurons located in L1 displayed strong activation by foot shocks (Fig. 1e). Layer 1 is unique in the neocortex because it contains only very few neuronal somata, almost all of which are GABAergic interneurons[22–25]. To investigate their activation in more detail, we used two-photon targeted loose-seal cell-attached recordings[26] (Fig. 1f). These experiments showed that L1 interneurons are tonically active (baseline frequency 3.2 ± 0.5 Hz, $n = 30$), and confirmed that the majority of L1 interneurons are strongly activated by foot shocks, while a minority (23%) displayed inhibition during and after foot shocks (Fig. 1g and Supplementary Fig. 3). The excitatory response was also present when foot shocks were paired with tones, whereas L1 interneurons did not show pronounced responses to tones alone (Supplementary Fig. 4). Interestingly, we observed a very similar activation of L1 interneurons by foot shocks in the primary visual cortex (Supplementary Fig. 5), indicating that the relevance of this pathway may extend beyond auditory memory acquisition.

Layer 1 is a prominent feedback pathway in the neocortex, containing both glutamatergic projections from higher cortical areas[27,28] (and from non-specific thalamic nuclei[29]) and cholinergic afferents from the basal forebrain, the major source of acetylcholine in the rodent neocortex[14,30–32]. We next sought to identify the afferent pathways mediating activation of L1 interneurons during foot shocks. Local bath-application of the glutamate receptor antagonist NBQX (1 mM) strongly reduced baseline firing frequency (Supplementary Fig. 6), but left the foot-shock response in L1 interneurons intact (Fig. 2a, b and Supplementary Fig. 6, see below). In contrast, combined application of the nicotinic acetylcholine receptor (nAChR) antagonists mecamylamine and methyllycaconitine (1 and 0.1 mM, respectively) abolished the L1 interneuron response almost completely (Fig. 2c and Supplementary Fig. 6), and drastically reduced baseline firing (Supplementary Fig. 6). In agreement with the interpretation that cholinergic basal forebrain afferents generate foot-shock responses in L1 interneurons, electrical microstimulation of the basal forebrain caused strong excitation of L1 interneurons in the absence of foot shocks (Fig. 2d and Supplementary Fig. 6). Activation of the basal forebrain by foot shocks is expected to have a longer latency than thalamocortical signalling[14]. Latency analysis showed that L1 interneuron activation was biphasic, with an early, glutamatergic peak (onset 10 to 20 ms after shock onset) which may originate from non-lemniscal thalamus[29], and a later, nicotinic peak (onset 40 to 50 ms after shock onset) that outlasted the foot shock and contained the majority of spikes (Fig. 2 insets). As previously reported[24], recordings from brain

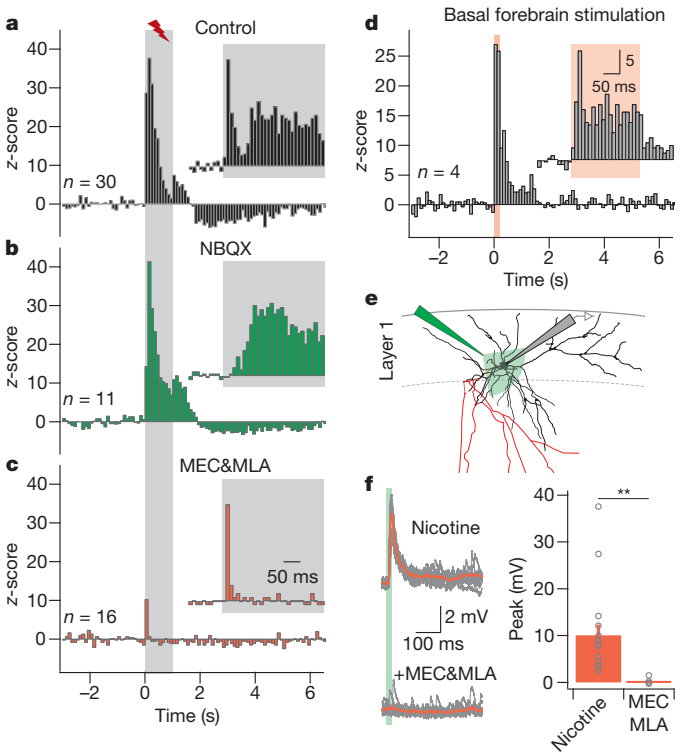


Figure 2 | Nicotinic activation of L1 interneurons by foot shocks. **a**, Population response of L1 interneurons to foot shocks. Insets in **a**–**d** are at high temporal resolution (10-ms bins). The control response was biphasic (onset 10–20 ms after foot-shock onset). **b**, Local application of the glutamate receptor antagonist NBQX left the response intact, but selectively eliminated the early peak, indicating its glutamatergic nature (inset, onset latency 40–50 ms). **c**, Local block of nAChRs by mecamylamine and methyllycaconitine (MEC&MLA) strongly reduced foot-shock responses, selectively eliminating the later, broader peak (inset). **d**, Electrical microstimulation of the basal forebrain activated L1 interneurons after 10 to 20 ms (inset, $n = 4$). **e**, Whole-cell recording of a L1 interneuron (black, soma and dendrites; red, axon) in auditory cortex slices during puff application of nicotine (100 µM, 20 ms). **f**, Left, example nicotine responses in control and after bath-application of mecamylamine (100 µM) and methyllycaconitine (0.1 µM; grey, single trials; red, average). Right, all L1 interneurons displayed nicotine responses ($n = 17$), which were blocked by the nAChR antagonists ($n = 8$). Values are mean ± s.e.m. $**P < 0.01$. Statistical analysis in Supplementary Information.

©2011 Macmillan Publishers Limited. All rights reserved

slices showed that all L1 interneurons showed responses to nicotine puffs (Fig. 2e, f) that were blocked by the same antagonists that abolished foot-shock responses *in vivo* (Fig. 2f), and could fire L1 interneurons (Supplementary Fig. 7). In summary, these results indicate that activity of cholinergic basal forebrain neurons is both necessary and sufficient to fire L1 interneurons during foot shocks, and that acetylcholine activates nAChRs on L1 interneurons. This implies that acetylcholine is released rapidly (<50 ms) after an aversive stimulus. Activation of L1 interneurons in turn is likely to have a central role in fear-conditioning-related plasticity in the cortex.

L1 interneurons inhibit L2/3 PV$^+$ interneurons

How do foot-shock responses in L1 interneurons affect processing in the local microcircuit? There is evidence that L1 interneurons can inhibit interneurons in L2/3 during nicotinic activation[24]. Fast-spiking, PV$^+$ interneurons are the most abundant type of interneuron in the neocortex[3]. We injected a conditional adeno-associated virus (AAV) expressing a fluorescent marker (venus) into auditory cortex of PV-ires-Cre mice[33] to label these cells selectively (Fig. 3a). Targeted recordings revealed that PV$^+$ interneurons in L2/3 are tonically active under baseline conditions (5.9 ± 1.2 Hz, $n = 17$), similar to fast-spiking interneurons in awake, head-fixed mice[34]. Foot shocks caused prominent, long-lasting inhibition of firing in the majority of PV$^+$ interneurons (88%, Fig. 3b and Supplementary Fig. 8), while the remaining two neurons displayed an excitatory response (Supplementary Fig. 9). Inhibition of PV$^+$ interneurons was strongly reduced by the nAChR antagonists mecamylamine and methyllycaconitine (Fig. 3b and Supplementary Fig. 8), which also blocked excitation of L1 interneurons by foot shocks (Fig. 2c). Because PV$^+$ fast-spiking interneurons lack intrinsic responses to acetylcholine[10,35], this result is consistent with direct inhibition of PV$^+$ interneurons by L1 interneurons. In line with this interpretation, we observed morphological and functional synaptic contacts between L1 interneurons and PV$^+$ interneurons (Supplementary Fig. 10). In addition, the time course of foot-shock responses in the two populations matches (Supplementary Fig. 11). Taken together, these data indicate that L2/3 fast-spiking PV$^+$ interneurons are inhibited by L1 interneurons during foot shocks.

To test whether this mechanism is also engaged in awake, freely moving animals, we implanted mice with single-unit recording electrodes in the superficial layers of the auditory cortex (Fig. 3c and Supplementary Fig. 12). Putative interneurons were distinguished from putative pyramidal neurons using unsupervised cluster analysis (Fig. 3d and Supplementary Fig. 12). Recordings during fear conditioning confirmed that a large population of putative interneurons is inhibited during and after an aversive shock (Fig. 3e and Supplementary Fig. 12). The same neurons were activated by the CS, indicating that the shock removes feed-forward inhibition in pyramidal neurons during auditory input. A similar proportion of putative interneurons displayed either excitation or no response to shocks (Supplementary Fig. 12). These data are consistent with the interpretation that excitation of L1 interneurons by aversive stimuli serves to remove both spontaneous and feed-forward inhibition provided by PV$^+$ interneurons to surrounding pyramidal neurons in behaving mice.

Disinhibition of L2/3 pyramidal neurons

PV$^+$ basket cells provide strong, perisomatic inhibition to local pyramidal neurons[3,6,7,10]. To test directly whether disinhibition is the main effect of foot shocks in auditory cortex L2/3 pyramidal cells, we used whole-cell recordings. Foot shocks elicited a depolarization from rest in all neurons tested (5 ± 1.1 mV, $n = 6$, Fig. 4a). The amplitude increased at more depolarized membrane potentials (Supplementary Fig. 13), consistent with a response mediated by strong disinhibition and weak excitation[36]. Application of nAChR antagonists converted the foot-shock response to a net inhibition (Fig. 4a), suggesting an involvement of the disinhibitory circuit described here, and perhaps indicating that its block unmasks a component of inhibition recruited by foot shocks. Finally, recordings under conditions isolating inhibitory postsynaptic currents (IPSCs)[17] revealed a drastic reduction

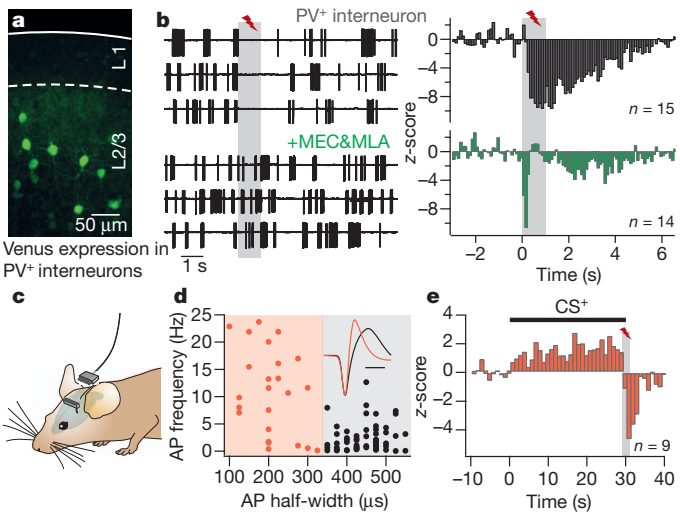


Figure 3 | Aversive shocks inhibit layer 2/3 PV$^+$ interneurons. **a**, Expression of venus in L2/3 PV$^+$ interneurons. **b**, Cell-attached recordings from PV$^+$ interneurons in control (top) show strong inhibition of spontaneous firing by foot shocks (left, example traces; right, population response). This response was strongly reduced in the presence of mecamylamine and methyllycaconitine (bottom). **c**, Single-unit recordings in the superficial layers of auditory cortex in freely behaving mice. **d**, Putative interneurons (red) were separated from putative pyramidal neurons (black) by unsupervised cluster analysis. Inset, example action potential (AP) waveforms (scale bar 400 µs). **e**, *z*-scored activity of putative interneurons showing inhibition by periorbital shocks during fear conditioning. Note recruitment by the CS$^+$ and subsequent inhibition by the shock ($n = 9$ of 24 putative interneurons, Supplementary Fig. 12).

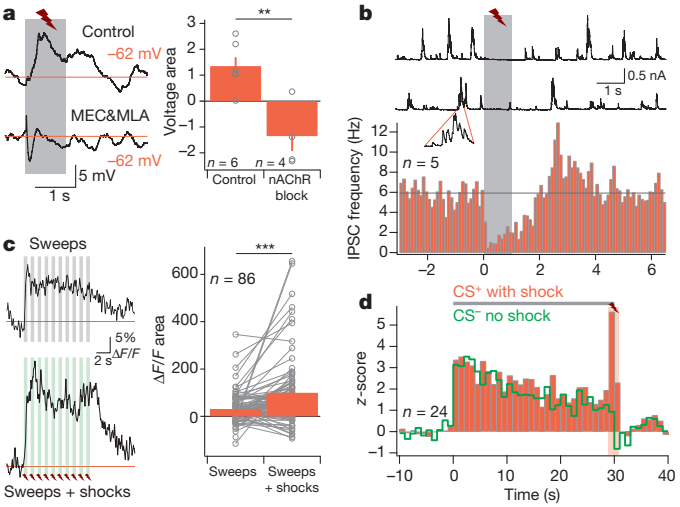


Figure 4 | Aversive shocks disinhibit L2/3 pyramidal neurons. **a**, Top left, *in vivo* whole-cell current-clamp recording of an auditory cortex pyramidal neuron during foot shocks. Bottom left, foot-shock response in the presence of mecamylamine and methyllycaconitine. Response area recorded at rest was reduced by nAChR antagonists (right). **b**, Whole-cell voltage-clamp recordings of inhibitory postsynaptic currents in pyramidal neurons show a strong reduction in IPSC frequency during and after foot shocks (top, example traces; bottom, population data, V_{hold} 0 to +20 mV). **c**, Left, example response of a L2/3 neuron to frequency-modulated sweeps (top) and sweeps paired with foot shocks (bottom) recorded with two-photon calcium imaging. Right, foot shocks caused a threefold enhancement of sweep response area in L2/3. **d**, Single-unit recordings of putative pyramidal neurons with significant response to the CS$^+$ in freely-behaving mice during fear conditioning. Note strong activation by coincidence of sweep and shock. Values are mean ± s.e.m. $^{**}P < 0.01$, $^{***}P < 0.001$. Statistical analysis in Supplementary Information.

©2011 Macmillan Publishers Limited. All rights reserved

of IPSC frequency during and after the foot shock (Fig. 4b). These results indicate that inhibition of PV^+ interneurons is a dominant influence shaping foot-shock responses in pyramidal neurons, and are in line with the observation that basal forebrain stimulation causes disinhibition of pyramidal cells[17].

Given that acetylcholine can affect many stages of neocortical information processing[8–11], we next aimed to determine how sensory input interacts with foot-shock-mediated disinhibition. We imaged calcium responses to tones in L2/3 of auditory cortex, and compared them to presentations of the same tones in conjunction with foot shocks (Fig. 4c and Supplementary Fig. 14). This experiment revealed that foot shocks cause a strong enhancement of the calcium signal integral, suggesting that tone/shock compounds elicit much greater activity than tones alone. A similar observation was made with extracellular recordings in freely moving mice, where coincidence of tone and shock excited putative pyramidal neurons much more than tone alone (Fig. 4d). This effect was highly supralinear in both experiments because shocks alone elicited almost no activity in these neurons (Supplementary Fig. 14). In summary, we provide evidence that L2/3 pyramidal neurons are disinhibited by aversive stimuli via inhibition of PV^+ interneurons.

Fear learning requires auditory cortex disinhibition

To determine whether activation of nAChRs contributes to fear learning, we applied mecamylamine and methyllycaconitine locally into the auditory cortex before fear conditioning (Fig. 5a and Supplementary Fig. 1). When tested 24 h later in drug-free state, this manipulation resulted in strongly reduced fear levels (Fig. 5b), consistent with a requirement for nicotinic activation of L1 interneurons. However, L1 interneurons are not the only circuit elements expressing nAChRs[8,9,11]. To test further whether disinhibition specifically during the foot shock contributes to fear learning, we used channelrhodopsin-2 (ChR-2)[37] expression in PV^+ interneurons (Fig. 5c and Supplementary Figs 15 and 16). Mice with chronic, bilateral optic fibre implantation (Fig. 5d) were subjected to differential fear conditioning in which optogenetic stimulation of PV^+ interneurons occurred during and for 5 s after the foot shock (Fig. 5e), the period during which we observed inhibition of these neurons (Fig. 3b). When tested 24 h later without optogenetic stimulation, these mice showed strongly reduced fear responses to the conditioned stimulus compared to sham-injected littermates (Fig. 5f). Reconditioning without optogenetic manipulation yielded normal fear learning (Fig. 5f). In addition, we ruled out the possibility that foot shock perception was perturbed by optogenetic stimulation and that laser illumination itself was perceived as a conditioned stimulus (Supplementary Fig. 17). Together, these data indicate that nicotinic disinhibition of the auditory cortex selectively during foot shock is required for associative fear learning.

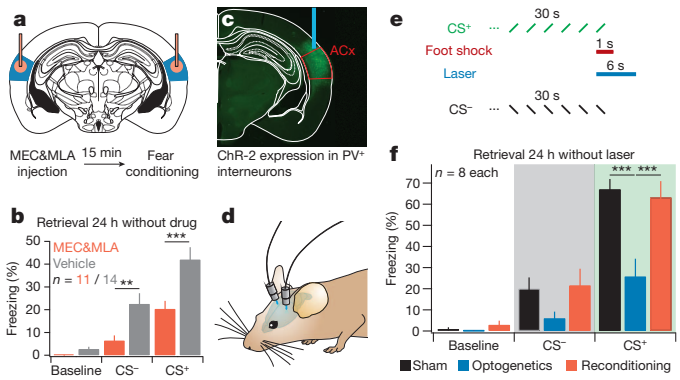


Figure 5 | Auditory cortex disinhibition is required for fear learning. **a**, Injection of mecamylamine and methyllycaconitine (red) into auditory cortex (blue) before fear conditioning. **b**, Drug-free freezing in a novel context one day after conditioning. Compared to vehicle-injected mice (grey), nAChR antagonism reduced freezing to both CS^- and CS^+ (red). **c**, Stimulation of ChR-2 expressing PV^+ interneurons (green) in auditory cortex (red) via an optic fibre (blue). **d**, Optogenetic manipulation in freely behaving mice. **e**, Differential fear conditioning protocol with optogenetic stimulation during and for 5 s after every foot shock. **f**, Freezing in a novel context without laser stimulation one day after conditioning. Compared to identically treated sham-injected littermates (black), virus-injected mice (blue) exhibit drastically reduced freezing to the CS^+. Reconditioning without optogenetic stimulation yielded strongly enhanced freezing (red) that was indistinguishable from sham. Values are mean ± s.e.m. $^{**}P < 0.01$, $^{***}P < 0.001$. Statistical analysis in Supplementary Information.

Discussion

Using targeted recordings from identified populations of auditory cortex neurons in conjunction with single-unit recordings, pharmacological and optogenetic manipulations in behaving mice, we have identified a disinhibitory microcircuit required for associative learning. Our data show that L1 interneurons play a central role in conveying information about an aversive stimulus to auditory cortex. Activity of L1 interneurons was tightly controlled by endogenous acetylcholine released from basal forebrain cholinergic projections, which determined baseline firing and acutely activated the majority of L1 interneurons during foot shocks, while a sub-set responded with inhibition. Given that all L1 interneurons express functional nAChRs[24] (Fig. 2), a likely source of this inhibition are synaptic interactions within L1 (ref. 25). Layer 1 contains several morphologically distinct subtypes of interneurons[22–25,38], and it remains to be determined whether foot-shock response type correlates with morphology. A strikingly similar activation by foot shocks was also observed in the primary visual cortex, which could underlie aspects of contextual fear learning and indicates that cholinergic activation of L1 interneurons may be a general feature of neocortical organization. Together, these results add to mounting evidence that L1 is a prominent locus of feedback signalling in the neocortex[27,28,32] and begin to define how feedback signals interact with thalamocortical feed-forward information during memory acquisition. Interestingly, L1 interneurons also receive prominent corticocortical feedback[27], and are responsive to dopamine[39] and serotonin[40], indicating that, depending on the nature of the learning task, different systems can feed into the microcircuit described here.

Foot shocks reduced both spontaneous and feed-forward perisomatic inhibition provided by basket cells, thus disinhibiting pyramidal neurons. Importantly, we cannot rule out that electrotonically remote dendritic sites received unchanged or even increased inhibition, for instance from L1 neurogliaform cells[25,38], which could serve to compartmentalize induction of synaptic plasticity. Nicotinic enhancement of inhibitory input to pyramidal neurons has been observed *in vitro*[11,41]. Therefore, the level of ongoing inhibition onto pyramidal neurons, as a prerequisite for disinhibition, may determine whether the net effect of nAChR activation is inhibitory or disinhibitory. Disinhibition is an attractive mechanism for learning because it is permissive for strong activation and concomitant plasticity induction, but not necessarily causative. Rather, the available sensory input at the time of the aversive stimulus can select the circuit elements for plasticity induction in a bottom-up fashion. In addition, other cholinergic actions like enhancement of thalamocortical transmission[11,42] and reduced GABA release from basket cell synapses[10] may have acted in synergy with the microcircuit described here to boost sensory responses.

Disinhibition of pyramidal neurons by foot shocks in turn probably gated the induction of activity-dependent plasticity in the auditory cortex[16,17] and at cortical afferents to the amygdala[18]. In parallel, cholinergic activation of L1 interneurons may also contribute to memory expression, because basal forebrain neurons acquire a conditioned response during learning[31,43]. Irrespective of the plasticity loci, our results delineate a role for the auditory cortex in fear conditioning to complex tones that goes beyond mere signalling of tone information to the amygdala[18–21]. Rather, we observe that stimulus conver-

©2011 Macmillan Publishers Limited. All rights reserved

gence and concomitant auditory cortex disinhibition are essential for fear learning. It is likely that the use of complex, naturalistic tones has emphasized the role of the auditory cortex[18], because fear conditioning to pure tones is often unaffected by auditory cortex lesions[19,21] (but see refs 19 and 20).

Beyond memory formation, cholinergic transmission is well known to mediate functions such as arousal and attention[30,31]. Interestingly, the transition from quiet wakefulness to active whisking is associated with a sharp drop in firing of fast-spiking interneurons in the barrel cortex[34], which is reminiscent of foot-shock responses in these neurons, and perhaps implies that cholinergic gating by inhibition of PV^+ interneurons is a general hallmark of active brain states.

METHODS SUMMARY

Male C57BL6/J mice (1.5 to 6 months old) were used. Full details of animals, materials and methods used for two-photon calcium imaging[44,45], patch-clamp[26] and single-unit recordings, pharmacological and optogenetic intervention[37] and behaviour[46] are provided in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 July; accepted 26 October 2011.
Published online 7 December 2011.

1. Martin, S. J. & Morris, R. G. New life in an old idea: the synaptic plasticity and memory hypothesis revisited. *Hippocampus* **12,** 609–636 (2002).
2. Hübener, M. & Bonhoeffer, T. Searching for engrams. *Neuron* **67,** 363–371 (2010).
3. Markram, H. *et al.* Interneurons of the neocortical inhibitory system. *Nature Rev. Neurosci.* **5,** 793–807 (2004).
4. Gabernet, L., Jadhav, S. P., Feldman, D. E., Carandini, M. & Scanziani, M. Somatosensory integration controlled by dynamic thalamocortical feed-forward inhibition. *Neuron* **48,** 315–327 (2005).
5. Wehr, M. & Zador, A. M. Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature* **426,** 442–446 (2003).
6. Freund, T. F. & Katona, I. Perisomatic inhibition. *Neuron* **56,** 33–42 (2007).
7. Kawaguchi, Y. & Kubota, Y. GABAergic cell subtypes and their synaptic connections in rat frontal cortex. *Cereb. Cortex* **7,** 476–486 (1997).
8. Lawrence, J. J. Cholinergic control of GABA release: emerging parallels between neocortex and hippocampus. *Trends Neurosci.* **31,** 317–327 (2008).
9. Bacci, A., Huguenard, J. R. & Prince, D. A. Modulation of neocortical interneurons: extrinsic influences and exercises in self-control. *Trends Neurosci.* **28,** 602–610 (2005).
10. Kruglikov, I. & Rudy, B. Perisomatic GABA release and thalamocortical integration onto neocortical excitatory cells are regulated by neuromodulators. *Neuron* **58,** 911–924 (2008).
11. Metherate, R. Nicotinic acetylcholine receptors in sensory cortex. *Learn. Mem.* **11,** 50–59 (2004).
12. Vogels, T. P. & Abbott, L. F. Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. *Nature Neurosci.* **12,** 483–491 (2009).
13. Suga, N. & Ma, X. Multiparametric corticofugal modulation and plasticity in the auditory system. *Nature Rev. Neurosci.* **4,** 783–794 (2003).
14. Weinberger, N. M. Auditory associative memory and representational plasticity in the primary auditory cortex. *Hear. Res.* **229,** 54–68 (2007).
15. Quirk, G. J., Armony, J. L. & LeDoux, J. E. Fear conditioning enhances different temporal components of tone-evoked spike trains in auditory cortex and lateral amygdala. *Neuron* **19,** 613–624 (1997).
16. Ji, W., Suga, N. & Gao, E. Effects of agonists and antagonists of NMDA and ACh receptors on plasticity of bat auditory system elicited by fear conditioning. *J. Neurophysiol.* **94,** 1199–1211 (2005).
17. Froemke, R. C., Merzenich, M. M. & Schreiner, C. E. A synaptic memory trace for cortical receptive field plasticity. *Nature* **450,** 425–429 (2007).
18. LeDoux, J. E. Emotion circuits in the brain. *Annu. Rev. Neurosci.* **23,** 155–184 (2000).
19. Campeau, S. & Davis, M. Involvement of subcortical and cortical afferents to the lateral nucleus of the amygdala in fear conditioning measured with fear-potentiated startle in rats trained concurrently with auditory and visual conditioned stimuli. *J. Neurosci.* **15,** 2312–2327 (1995).
20. Boatman, J. A. & Kim, J. J. A thalamo-cortico-amygdala pathway mediates auditory fear conditioning in the intact brain. *Eur. J. Neurosci.* **24,** 894–900 (2006).
21. Romanski, L. M. & LeDoux, J. E. Bilateral destruction of neocortical and perirhinal projection targets of the acoustic thalamus does not disrupt auditory fear conditioning. *Neurosci. Lett.* **142,** 228–232 (1992).
22. Zhou, F. M. & Hablitz, J. J. Morphological properties of intracellularly labeled layer I neurons in rat neocortex. *J. Comp. Neurol.* **376,** 198–213 (1996).
23. Hestrin, S. & Armstrong, W. E. Morphology and physiology of cortical neurons in layer I. *J. Neurosci.* **16,** 5290–5300 (1996).
24. Christophe, E. *et al.* Two types of nicotinic receptors mediate an excitation of neocortical layer I interneurons. *J. Neurophysiol.* **88,** 1318–1327 (2002).
25. Chu, Z., Galarreta, M. & Hestrin, S. Synaptic interactions of late-spiking neocortical neurons in layer 1. *J. Neurosci.* **23,** 96–102 (2003).
26. Margrie, T. W. *et al.* Targeted whole-cell recordings in the mammalian brain *in vivo*. *Neuron* **39,** 911–918 (2003).
27. Gonchar, Y. & Burkhalter, A. Distinct GABAergic targets of feedforward and feedback connections between lower and higher areas of rat visual cortex. *J. Neurosci.* **23,** 10904–10912 (2003).
28. Cauller, L. J., Clancy, B. & Connors, B. W. Backward cortical projections to primary somatosensory cortex in rats extend long horizontal axons in layer I. *J. Comp. Neurol.* **390,** 297–310 (1998).
29. Rubio-Garrido, P., Perez-de-Manzo, F., Porrero, C., Galazo, M. J. & Clasca, F. Thalamic input to distal apical dendrites in neocortical layer 1 is massive and highly convergent. *Cereb. Cortex* **19,** 2380–2395 (2009).
30. Hasselmo, M. E. & Sarter, M. Modes and models of forebrain cholinergic neuromodulation of cognition. *Neuropsychopharmacology* **36,** 52–73 (2011).
31. Wenk, G. L. The nucleus basalis magnocellularis cholinergic system: one hundred years of progress. *Neurobiol. Learn. Mem.* **67,** 85–95 (1997).
32. Mechawar, N., Cozzari, C. & Descarries, L. Cholinergic innervation in adult rat cerebral cortex: a quantitative immunocytochemical description. *J. Comp. Neurol.* **428,** 305–318 (2000).
33. Hippenmeyer, S. *et al.* A developmental switch in the response of DRG neurons to ETS transcription factor signaling. *PLoS Biol.* **3,** e159 (2005).
34. Gentet, L. J., Avermann, M., Matyas, F., Staiger, J. F. & Petersen, C. C. Membrane potential dynamics of GABAergic neurons in the barrel cortex of behaving mice. *Neuron* **65,** 422–435 (2010).
35. Kawaguchi, Y. Selective cholinergic modulation of cortical GABAergic cell subtypes. *J. Neurophysiol.* **78,** 1743–1747 (1997).
36. Manookin, M. B., Beaudoin, D. L., Ernst, Z. R., Flagel, L. J. & Demb, J. B. Disinhibition combines with excitation to extend the operating range of the OFF visual pathway in daylight. *J. Neurosci.* **28,** 4136–4150 (2008).
37. Zhang, F., Aravanis, A. M., Adamantidis, A., de Lecea, L. & Deisseroth, K. Circuit-breakers: optical technologies for probing neural signals and systems. *Nature Rev. Neurosci.* **8,** 577–581 (2007).
38. Wozny, C. & Williams, S. R. Specificity of synaptic connectivity between layer 1 inhibitory interneurons and layer 2/3 pyramidal neurons in the rat neocortex. *Cereb. Cortex* **21,** 1818–1826 (2011).
39. Wu, J. & Hablitz, J. J. Cooperative activation of D1 and D2 dopamine receptors enhances a hyperpolarization-activated inward current in layer I interneurons. *J. Neurosci.* **25,** 6322–6328 (2005).
40. Foehring, R. C., van Brederode, J. F., Kinney, G. A. & Spain, W. J. Serotonergic modulation of supragranular neurons in rat sensorimotor cortex. *J. Neurosci.* **22,** 8238–8250 (2002).
41. Couey, J. J. *et al.* Distributed network actions by nicotine increase the threshold for spike-timing-dependent plasticity in prefrontal cortex. *Neuron* **54,** 73–87 (2007).
42. Gil, Z., Connors, B. W. & Amitai, Y. Differential regulation of neocortical synapses by neuromodulators and activity. *Neuron* **19,** 679–686 (1997).
43. Acquas, E., Wilson, C. & Fibiger, H. C. Conditioned and unconditioned stimuli increase frontal cortical and hippocampal acetylcholine release: effects of novelty, habituation, and fear. *J. Neurosci.* **16,** 3089–3096 (1996).
44. Stosiek, C., Garaschuk, O., Holthoff, K. & Konnerth, A. *In vivo* two-photon calcium imaging of neuronal networks. *Proc. Natl Acad. Sci. USA* **100,** 7319–7324 (2003).
45. Nimmerjahn, A., Kirchhoff, F., Kerr, J. N. & Helmchen, F. Sulforhodamine 101 as a specific marker of astroglia in the neocortex *in vivo*. *Nature Methods* **1,** 31–37 (2004).
46. Ciocchi, S. *et al.* Encoding of conditioned fear in central amygdala inhibitory circuits. *Nature* **468,** 277–282 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank all members of the Lüthi laboratory for discussions and comments. We thank B. Kampa and T. Oertner for advice on two-photon imaging, R. Friedrich and J. Gründemann for comments on the manuscript, P. Argast, J. Lüdke and C. Müller for excellent technical assistance, H. Zielinska for preparation of artwork and S. Arber, Y. Yanagawa and K. Deisseroth for generously sharing materials. This work was supported by grants from the Swiss National Science Foundation (J.J.L. Ambizione; A.L.), the National Competence Center in Research (NCCR) of the Swiss Confederation on the synaptic basis of mental disorders, the French National Research Agency (C.H., ANR-2010-BLAN-1442-01), a Marie-Curie fellowship (J.J.L.), a Schering Foundation fellowship (S.B.E.W.), a Fonds AXA pour la Recherche fellowship (J.C.), the Aquitaine Regional Council (C.H.) and the Novartis Research Foundation.

Author Contributions J.J.L. initiated the project and performed most experiments and data analyses. S.B.E.W. established optogenetic manipulation. S.B.E.W. and P.T. helped with optogenetic experiments. E.M.M.M. performed and analysed *in vitro* experiments. P.T. performed and analysed *in vivo* pharmacology. J.C. and C.H. performed and analysed single-unit recordings. J.J.L. and A.L. conceived the project and wrote the manuscript. All authors contributed to the experimental design and interpretation, and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.J.L. (johannes.letzkus@fmi.ch) or A.L. (andreas.luthi@fmi.ch).

©2011 Macmillan Publishers Limited. All rights reserved

METHODS

Animals. Male C57BL6/J mice (1.5–6 months old, Harlan and Janvier) were housed under a 12 h light/dark cycle, and provided with food and water ad libitum. Prior to fear conditioning, mice were individually housed for ≥ 7 days and habituated to handling ≥ 5 times. All animal procedures were executed in accordance with institutional guidelines and the prescribed authorities (Veterinary Department of the Canton of Basel-Stadt; Switzerland or Ministère de l'agriculture et de la forêt, 87-848, France and European Economic Community, EEC, 86-6091).

Head-fixed two-photon imaging and targeted patch-clamp recordings. Mice were sedated with chlorprothixene (5 mg kg^{-1}, intraperitoneal injection), kept warm on a hot-water bottle and after 10 min anaesthetized with urethane (0.9 g per kg, intraperitoneal injection). Core body temperature was maintained at 37.5 °C by a feedback-controlled heating pad (FHC). Mice were fixed in a stereotaxic frame (Kopf Instruments), and local analgesia was provided by injection of ropivacain under the scalp (10 mg ml^{-1}, Naropin, AstraZeneca). After retraction of the scalp, the location of the area of interest was determined as follows. Auditory cortex: 2.46 mm posterior of bregma, 4.5 to 4.7 mm lateral of midline; primary visual cortex: 3.16 mm posterior of bregma, 2.5 mm lateral of midline. A custom-built head-plate was glued to the skull with cyanoacrylate glue (Ultra Gel, Henkel) and dental cement (Paladur, Heraeus), and a small craniotomy was performed (2–3-mm diameter). The dura was retracted, the exposed cortical surface was superfused with normal rat ringer (NRR) containing (in mM): 135 NaCl, 5.4 KCl, 5 HEPES, 1.8 $CaCl_2$, pH 7.2 with NaOH), covered with agarose (1% type-III-A in NRR) and fixed with a coverslip. For two-photon calcium imaging, the membrane permeant dye Oregon Green BAPTA-1 acetoxymethyl ester (OGB-1, 1 mM in NRR with 8% DMSO and 2% pluronic F-127, Invitrogen) was pressure-ejected from a patch pipette under two-photon visualization. After 1 h, this lead to intracellular staining of virtually all cells in a diameter of approximately 300 μm[44]. Localization of the labelled area in auditory cortex was verified post hoc in slices. Sulforhodamine 101 (0.2 mM in NRR) was bath-applied for 5 min to the exposed cortex to counterstain glial cells[45]. Two-photon calcium imaging was performed using a scanning microscope (Fluoview 1000MPE, Olympus) coupled to a femtosecond-pulse infrared laser at 830 nm wavelength (MaiTai HP, Spectra-Physics). The beam was adjusted with a telescope to fill the back-focal plane of the ×20 water-immersion objective (0.95 numerical aperture, Olympus). Average power in the back-focal plane was <120 mW. Frames were acquired (Fluoview software, Olympus) at 15 Hz at depths below the pia ranging from 20 to 400 μm. Time stacks were processed offline using ImageJ to extract $\Delta F/F$ for each neuron. Foot shocks (1 s, direct current, 0.4 to 2 mA, 10–20 s interval) were delivered to the hindpaws. Frequency-modulated sweeps (logarithmically modulated between 5 and 15 kHz, 50 ms rise and fall) were generated (System3, Tucker Davis Technologies) and presented at 70 dB sound pressure level from an electrostatic speaker (ES1, TDT) mounted in front of the animal.

Two-photon-guided patch-clamp recordings were performed using patch pipettes of 4–10 MΩ resistance filled with NRR for cell-attached recordings, for whole-cell current-clamp recordings with intracellular solution containing (in mM): 130 K-methanesulphonate, 6.3 KCl, 20 Na_2-phosphocreatine, 4 Mg-ATP, 0.3 Na-GTP, 10 HEPES, 290 mOsm, pH 7.3 with KOH and for whole-cell voltage-clamp recordings with intracellular solution containing (in mM): 125 Cs-methanesulphonate, 2 CsCl, 10 Na_2-phosphocreatine, 4 Mg-ATP, 0.3 Na-GTP, 5 TEACl, 3.5 QX-314, 0.5 EGTA, 10 HEPES, 290 mOsm, pH 7.3 with CsOH. Pipette solutions were supplemented with Alexa 488 or Alexa 594 (25–50 μM, Invitrogen). Neurons of interest were identified by either marker expression (L1 interneurons in GAD67–GFP mice, PV^+ interneurons in PV-ires-Cre mice[33] injected with AAV conditionally expressing venus), or by imaging the 'shadows' created by neurons in dye-filled extracellular space in wild-type mice (L1 interneurons, L2/3 pyramidal neurons). L1 interneurons were located between 30 and 70 μm below the pia. Cell-attached and whole-cell recordings were established using a ×40 water-immersion objective (0.8 numerical aperture, Olympus) according to standard procedures[26]. Cell-attached recordings were performed in current-clamp or voltage-clamp configuration, and rejected when no action potential occurred during the entire length of the recording (≥5 min). Whole-cell current-clamp recordings were rejected when the series resistance exceeded 70 MΩ. Whole-cell voltage-clamp recordings were performed at 0–20 mV holding potential and rejected when the series resistance exceeded 50 MΩ. IPSCs were detected using a derivative threshold of 40 pA ms^{-1}. Pyramidal neurons were identified after dye filling by pyramidal morphology and spiny dendrites. Signals were recorded (Axopatch 200B, Molecular Devices or ELC-03XS, NPI), low-pass filtered at 20 kHz and digitized at 50 kHz (Digidata 1322A, Molecular Devices) using pClamp software (Molecular Devices). For cell-attached recordings, action potentials were detected based on amplitude. Peri-stimulus time histograms were computed across the entire population of recordings of a given type. Microstimulation of the basal forebrain (20 pulses of 0.2 ms duration delivered at 100 Hz) was performed using bipolar stimulating electrodes at the following coordinates: 0.8 mm posterior of bregma, 1.75 mm lateral of midline, 4.3 mm below cortical surface. Drugs were present in the bath solution throughout the experiment. Unless stated otherwise, chemicals were obtained from Sigma.

Surgery and local drug injection. Mice were anaesthetized with isoflurane (induction, 5%; maintenance, 1.5%) in oxygen-enriched air (Oxymat 3, Weinmann) and fixed in a stereotaxic frame (Kopf Instruments). Core body temperature was maintained at 36.5 °C by a feedback-controlled heating pad (FHC). Analgesia was provided by local injection of ropivacain under the scalp (Naropin, AstraZeneca) and systemic injection of meloxicam (100 μl of 5 mg ml^{-1}, intraperitoneal, Metacam, Boehringer-Ingelheim). Guide cannulae (26 gauge, with dummy screw caps, Plastics One) were implanted bilaterally to inject at the following coordinates: 2.46 mm posterior of bregma, ± 4.5 mm lateral of midline, 0.6 mm below cortical surface. Implants were fixed to the skull with cyanoacrylate glue (Ultra Gel, Henkel) and dental cement (Paladur, Heraeus). Mice were then given 1 week to recover from surgery, during which time they were handled ≥ 5 times to habituate them to the injection procedure. Fifteen minutes before fear conditioning, 32-gauge stainless steel injectors attached to 25 μl Hamilton syringes were inserted into the guide cannulae and an injection volume of 0.2 μl per hemisphere was delivered within 40 s using a microinfusion pump (Stoelting). Drug animals received bilateral injections of muscimol (100 ng per hemisphere) or a combination of mecamylamine (10 μg) and methyllycaconitine (10 μg) dissolved in NRR containing 1% fast green (Serva). Control mice were injected with vehicle solution only. In a subset of mice, fluorescent muscimol bodipy (625 μM in NRR with 5% DMSO) was injected after fear memory retrieval to quantify spread of the drug. After completion of the experiment, mice were transcardially perfused with 4% paraformaldehyde in phosphate-buffered saline (PFA), their brains extracted and post-fixed in paraformaldehyde overnight. For histological verification of the injection site, 60-μm coronal brain sections were made on a vibratome (Leica Microsystems) and imaged on a stereoscope (Leica Microsystems).

Virus injection and optogenetics. For labelling of PV^+ interneurons (Fig. 3a, b), an adeno-associated virus (AAV, serotype 2/7 (venus) or 2/9 (tdTomato), Vector Core, University of Pennsylvania) was used to deliver a DNA construct for conditional, Cre-dependent expression of venus or tdTomato. AAV (approximately 0.5 μl per hemisphere) was injected from glass pipettes (tip diameter 10–20 μm) connected to a picospritzer (Parker Hannifin Corporation) into the auditory cortex of PV-ires-Cre mice[33] at the following coordinates: 2.46 mm posterior of bregma, 4.5 mm lateral of midline, 0 to 1.1 mm below cortical surface. Experiments were performed after 0.5–3 months of expression time. For optical control of PV^+ interneurons (Fig. 5c–f), the same methodology was used for bilateral injection of a conditional AAV coexpressing channelrhodopsin-2 (ChR-2)[37] and venus (AAV2/7 EF1a::DIO-ChR2(H134R)-2A-NpHR-2A-Venus[47]) into the auditory cortex of PV-ires-Cre mice. These animals were additionally implanted with custom-built connectors holding optic fibres (0.48 numerical aperture, 200-μm diameter, Thorlabs). Fibre ends were inserted 200 μm into neocortex at the injection site. Implants were fixed to the skull with cyanoacrylate glue (Ultra Gel, Henkel) and dental cement (Paladur, Heraeus). After 2–4 months of expression time and habituation to handling, both implanted connectors were linked to a custom-built laser bench via optic fibres suspended over the conditioning context. This arrangement allowed the animals to move freely in the context. Stimulation of ChR-2 (<20 mW per implanted fibre) was delivered during and for 5 s after each foot shock (Fig. 5e) using a custom-built laser bench (laser: MBL473, 473-nm wavelength, CNI Lasers). One day after conditioning, mice were exposed to CS^+ and CS^- without optogenetic intervention in a neutral context. The optogenetic approach was previously validated *in vitro*[47] and *in vivo*[46], and here by acute extracellular recordings from auditory cortex of infected animals (Supplementary Fig. 16). Custom-built optrodes (optic fibre with 16-wire electrode attached) were lowered into the infected area. Laser pulses (300 ms) were delivered and recordings performed as detailed in section 'Extracellular recordings in freely behaving mice'. After completion of the experiment, mice were transcardially perfused with 4% paraformaldehyde. Brains were post-fixed in paraformaldehyde overnight at 4 °C, and cut into 80-μm thick coronal slices on a vibratome (Leica Microsystems). To improve the fluorescent signal, an immunostaining was performed. Slices were kept in blocking solution (3% BSA, 0.2% Triton in 0.1 M PBS) for 1 h at room temperature, before application of the primary antibody (Goat anti-GFP, Abcam; 1:500 in blocking solution) and incubated at 4 °C overnight. After washing, slices were incubated with secondary antibody (Alexa Fluor 488, donkey anti goat, Invitrogen; 1:1,000 in PBS with 3% BSA) at 4 °C overnight. After a final wash, slices were mounted on coverslips and imaged.

©2011 Macmillan Publishers Limited. All rights reserved

Behaviour. Fear conditioning and fear retrieval took place in two different contexts (context A and B). The conditioning and test boxes and the floor were cleaned before and after each session with 70% ethanol or 1% acetic acid, respectively. To score freezing behaviour, an automatic infrared beam detection system placed on the bottom of the experimental chambers (Coulbourn Instruments) was used. Mice were considered to be freezing if no movement was detected for 2 s and the measure was expressed as a percentage of time spent freezing. To ensure that our automatic system scores freezing rather than just immobility, we previously compared the values obtained with those measured using a classical time-sampling procedure during which an experimenter blind to the experimental conditions determined the mice to be freezing or not freezing every 2 s (defined as the complete absence of movement except for respiratory movements). The values obtained were 95% identical and the automatic detection system was therefore used throughout the experimental sessions. Conditioned stimuli for differential fear conditioning were 30-s trains of frequency-modulated sweeps (500 ms duration, logarithmically modulated between 5 and 15 kHz, 50 ms rise and fall) delivered at 1 Hz at a sound pressure level of 70 dB. The CS^+ (upsweep) was paired with a foot shock (1 s, 0.6 mA, 15 CS^+–foot-shock pairings; inter-trial interval: 20–180 s). The onset of the foot shock coincided with the onset of the last sweep in the CS^+. The CS^- (downsweep) was presented after each CS^+–foot-shock association, but was never reinforced (15 CS^- presentations, inter-trial interval: 20–180 s). On the next day, conditioned mice were submitted to fear retrieval in context B, during which they received 4 and 4 presentations of the CS^- and the CS^+, respectively.

Extracellular recordings in freely behaving mice. Surgical procedures are described earlier. Mice were secured in a stereotaxic frame and implanted with a pair of insulated silver wires (170-μm diameter) beneath the skin of each eyelid for delivery of periorbital shocks. In addition, mice were unilaterally implanted in the auditory cortex with a multi-wire electrode aimed at the following coordinates: 2.46 mm posterior of bregma, 4.5 mm lateral to midline, and 0.6 mm to 0.85 mm below the cortical surface. Electrodes consisted of 16 individually insulated, gold-plated nichrome wires (13 μm inner diameter, impedance 30 to 100 kΩ, Sandvik) contained in a 26-gauge stainless steel guide cannula. The wires were attached to a connector (18 pin, Omnetics). The implant was secured using cyanoacrylate adhesive gel. After surgery mice were allowed to recover for 7 days. Analgesia was applied before, and during the 3 days after surgery (Metacam, Boehringer-Ingelheim). Electrodes were connected to a headstage (Plexon) containing 16 unity-gain operational amplifiers. The headstage was connected to a 16-channel computer-controlled preamplifier (gain 100×, bandpass filter from 150 Hz to 9 kHz, Plexon). Neuronal activity was digitized at 40 kHz, bandpass filtered from 250 Hz to 8 kHz, and isolated by time–amplitude window discrimination and template matching using a multichannel acquisition processor system (Plexon). Single-unit spike sorting was performed using an off-line spike sorter (OFSS, Plexon). Principal component scores were calculated for unsorted waveforms and plotted on three-dimensional principal component spaces, and clusters containing similar valid waveforms were manually defined. A group of waveforms was considered to originate from a single neuron if it defined a discrete cluster in principal component space that was distinct from clusters for other units, and if it displayed a clear refractory period (1 ms) in auto-correlograms. Template waveforms were then calculated for well-separated clusters and stored for further analysis. To avoid analysis of the same neuron recorded on different channels, we computed cross-correlation histograms. If a target neuron displayed a peak of activity at a time that the reference neuron fired, only one of the two neurons was considered for further analysis. To separate putative inhibitory interneurons from putative excitatory pyramidal neurons we used an unsupervised cluster algorithm based on the Ward's method. Briefly, the Euclidian distance was calculated between all cell pairs based on the two-dimensional space defined by each cell's average spike width (measured from trough to peak) and baseline firing rate. An iterative agglomerative procedure was then used to combine cells into groups based on the matrix of distances such that the total number of groups was reduced to give the smallest possible increase in the within-group sum of square deviation. Mice were presented with the same conditioned stimuli used for fear conditioning (see above). The CS^+ was paired with a periorbital shock (2 s, 2.5 mA, 15 CS^+–shock pairings, inter-trial interval 20–180 s), which was used instead of foot shocks to minimize electrical artefacts in the recording. The onset of the periorbital shock coincided with the onset of the last sweep. The CS^- was presented after each CS^+ (15 CS^- presentations, inter-trial interval: 20–180 s). CS^+ and CS^- were counterbalanced across animals. At the conclusion of the experiment, recording sites were marked with electrolytic lesions before perfusion, and electrode locations were reconstructed with standard histological techniques

***In vitro* electrophysiology.** Coronal sections (300-μm thick) of auditory cortex were prepared from mice (6–8 weeks old) in ice-cold slicing artificial cerebrospinal fluid (ACSF) containing (in mM): 124 NaCl, 1.25 NaH_2PO_4, 10 $MgSO_4$, 2.7 KCl, 26 $NaHCO_3$, 2 $CaCl_2$, 10 Glucose, 4 ascorbate (95% O_2/5% CO_2). Slices were incubated for 45 min at 37 °C in an interface chamber, and then allowed to cool to room temperature. Recordings were performed at 34 °C under infrared videomicroscopy in ACSF containing (in mM): 124 NaCl, 1.25 NaH_2PO_4, 1.3 $MgSO_4$, 2.7 KCl, 26 $NaHCO_3$, 2 $CaCl_2$, 10 glucose, 4 ascorbate (95% O_2/5% CO_2). Whole-cell current-clamp recordings were performed with patch pipettes (4–10 MΩ resistance) filled with intracellular solution containing (in mM): 130 methanesulphonate, 6.3 KCl, 20 Na_2-phosphocreatine, 0.3 Na-GTP, 4 Mg-ATP, 10 HEPES, 5 biocytin, 290 mOsm, pH 7.3 with KOH. Signals were recorded (Multiclamp 700B, Molecular Devices), low-pass filtered at 10 kHz and digitized at 20 kHz (Digidata 1322A, Molecular Devices) using pClamp9 software (Molecular Devices). Signals were analysed using Igor Pro (Wavemetrics). The border between L1 and L2/3 was visually identified by the abrupt change in cell number, and only L1 interneurons at least 15 μm away from the border region were chosen for recordings. Local puff application of nicotine (100 μM, 20–80 ms duration) was delivered from a patch pipette connected to a picospritzer (Parker Hannifin Corporation). For paired recordings between L1 interneurons and L2/3 PV^+ interneurons, PV-ires-Cre mice were injected with an AAV (Vector Core, University of Pennsylvania) leading to Cre-dependent expression of tdTomato. Fluorescent PV^+ interneurons were identified using TillvisION (Till Photonics) and double recordings were performed as described above. Unless stated otherwise, drugs were bath-applied. To reveal morphology, slices were fixed in 4% paraformaldehyde (4 °C overnight), washed and stained for biocytin (4 days at 4 °C in PBS, 0.5–1% Triton-X, 0.2% Alexa (488, 568 or 680) streptavidin conjugate, Invitrogen). To identify putative synaptic contacts, this was combined with immunostaining against venus (see above) or tdTomato (Rabbit anti-RFP, MBL, Nunningen; 1:500 in blocking solution; Alexa Fluor 568, donkey anti-rabbit, Invitrogen; 1:1,000 in PBS with 3% BSA). Slices were mounted and confocal images acquired using an LSM700 confocal laser scanning microscope (Zeiss). Stacks were analysed using ImageJ (NIH) and single representative cells were reconstructed using Neurolucida software (MBF Bioscience). Putative synaptic contacts were scored when the axon of a L1 interneuron was located within 1 μm from a PV^+ dendrite or soma.

47. Tang, W. *et al.* Faithful expression of multiple proteins via 2A-peptide self-processing: a versatile and reliable method for manipulating brain circuits. *J. Neurosci.* **29,** 8621–8629 (2009).

©2011 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature10694

Interactions between cancer stem cells and their niche govern metastatic colonization

Ilaria Malanchi[1]*, Albert Santamaria-Martínez[1]*, Evelyn Susanto[1], Hong Peng[1,2], Hans-Anton Lehr[3], Jean-Francois Delaloye[4] & Joerg Huelsken[1]

Metastatic growth in distant organs is the major cause of cancer mortality. The development of metastasis is a multistage process with several rate-limiting steps[1]. Although dissemination of tumour cells seems to be an early and frequent event[2], the successful initiation of metastatic growth, a process termed 'metastatic colonization', is inefficient for many cancer types and is accomplished only by a minority of cancer cells that reach distant sites[3,4]. Prevalent target sites are characteristic of many tumour entities[5], suggesting that inadequate support by distant tissues contributes to the inefficiency of the metastatic process. Here we show that a small population of cancer stem cells is critical for metastatic colonization, that is, the initial expansion of cancer cells at the secondary site, and that stromal niche signals are crucial to this expansion process. We find that periostin (POSTN), a component of the extracellular matrix, is expressed by fibroblasts in the normal tissue and in the stroma of the primary tumour. Infiltrating tumour cells need to induce stromal POSTN expression in the secondary target organ (in this case lung) to initiate colonization. POSTN is required to allow cancer stem cell maintenance, and blocking its function prevents metastasis. POSTN recruits Wnt ligands and thereby increases Wnt signalling in cancer stem cells. We suggest that the education of stromal cells by infiltrating tumour cells is an important step in metastatic colonization and that preventing *de novo* niche formation may be a novel strategy for the treatment of metastatic disease.

We aimed to explore limiting factors that determine metastatic success using the MMTV-PyMT mouse breast cancer model, which spontaneously metastasizes to the lungs[6]. We reasoned that the recently identified cancer stem cells (CSCs, also called tumour-initiating cells), a subset of cancer cells that allow long-term tumour growth and are thought to be responsible for remissions[7,8], might also be relevant to the development of metastatic disease (Supplementary Fig. 1). We measured the relative size of the population of CSCs from primary MMTV-PyMT tumours and their pulmonary metastases using the previously established markers CD90 and CD24, which label a subset of the $CD24^{+}CD29^{hi}$ or $CD24^{+}CD49f^{hi}$ population used earlier to isolate CSCs and normal mammary gland stem cells[9–13] (Supplementary Fig. 2). This CSC subset accounts for $3 \pm 2.1\%$ (s.d.) of all tumour cells from both primary tumours and metastases (Fig. 1a). When $CD90^{+}CD24^{+}$ CSCs or $CD90^{+}CD24^{+}$-depleted non-CSCs are separately isolated from GFP^{+} tumours and directly introduced into mice through tail vein injection (GFP, green fluorescent protein), only the CSC population is able to produce lung metastases (Fig. 1b). Moreover, $CD90^{+}CD24^{+}$ cells isolated subsequently from pulmonary metastases are again the only tumour cell population that efficiently initiates secondary metastases (Fig. 1c). This is not due to differences in the extravasation capabilities of CSCs and non-CSCs (Supplementary Fig. 3).

In time course experiments, the relative size of the CSC population changes drastically during metastatic colonization. When injecting unfractionated tumour cells, the total number of tumour cells in the lung declines rapidly within the first 7 d after seeding[14] (Fig. 1d and Supplementary Fig. 4). Notably, the relative amount of CSCs transiently increases within the first and second weeks to more than 20%. This is due to selective expansion of the stem cell population as a result of their proliferation rate increasing over sixfold relative to the primary tumour in this early phase of metastatic colonization (as measured by BrdU incorporation; Fig. 1e and Supplementary Fig. 5). By contrast, non-CSCs show reduced proliferation 7 d after injection and, notably, fail to generate CSCs (Fig. 1e and Supplementary Fig. 6). Consequently, only CSCs are able to form metastatic colonies, and non-CSCs remain as single cells (Fig. 1f). Together, this demonstrates that selective expansion of CSCs is responsible for the initiation of metastasis.

The number of injected CSCs evidently exceeds the number of metastatic nodules, indicating that additional factors restrict successful metastatic colonization. Stem cells are suspected to rely on signals from their stromal environment, such as localized growth factors that can affect stem cell maintenance and proliferation[15–17]. We identified *Postn* as a stromal factor of normal stem cell niches and the metastatic niche (Supplementary Methods). *Postn* encodes for periostin, which becomes incorporated in the extracellular matrix[18] and has a role in bone, tooth and heart development and function[19–21]. Its expression is downregulated in the adult except in niches in direct contact with tissue-specific stem cells in mammary gland, bone, skin and intestine (Fig. 2a and Supplementary Figs 7 and 8). In tumours, POSTN is produced by stromal $\alpha SMA^{+}VIM^{+}$ (αSMA also known as ACTA2) fibroblasts according to RNA *in situ* hybridization, immunostaining and quantitative PCR (Fig. 2b–h). Notably, POSTN expression is induced in the lung stroma by infiltrating cancer cells (Fig. 2d), but does not occur in the alveolar lung tissue of tumour-bearing but metastasis-free animals (Supplementary Fig. 8). Human breast cancer patients show induction of stromal POSTN expression in 75% of lymph node metastases (Fig. 2i, j and Supplementary Fig. 9). Whereas POSTN-deficient mice maintain normal mammary gland development (Supplementary Figs 10–12), MMTV-PyMT $Postn^{-/-}$ breast cancers show a dramatic decrease in the number and size of pulmonary metastases, to less than 10% of controls, despite unaltered primary tumour size and morphology ($n = 37$, $P < 0.003$; Fig. 3a–d and Supplementary Figs 13–15). Metastasis formation from POSTN-deficient tumour cells is rescued in wild-type recipients (Supplementary Figs 16 and 17), where expression of POSTN is at the same level as in controls (Supplementary Fig. 18), indicating that stromal production of POSTN determines metastatic efficiency.

In searching for tumour-derived factors that can induce stromal POSTN, *in silico* promoter analysis predicted regulation mediated

[1]Ecole Polytechnique Fédérale de Lausanne, Swiss Institute for Experimental Cancer Research and National Center of Competence in Research 'Molecular Oncology', 1015 Lausanne, Switzerland. [2]Nanfang Hospital, Department of Otorhinolaryngology, Head and Neck Surgery, Southern Medical University, Guangzhou 510515, China. [3]University Institute of Pathology, CHUV, University of Lausanne, 1011 Lausanne, Switzerland. [4]Department of Gynecology and Obstetrics, Centre Hospitalier Universitaire Vaudois, 1011 Lausanne, Switzerland.
*These authors contributed equally to this work.

©2011 Macmillan Publishers Limited. All rights reserved

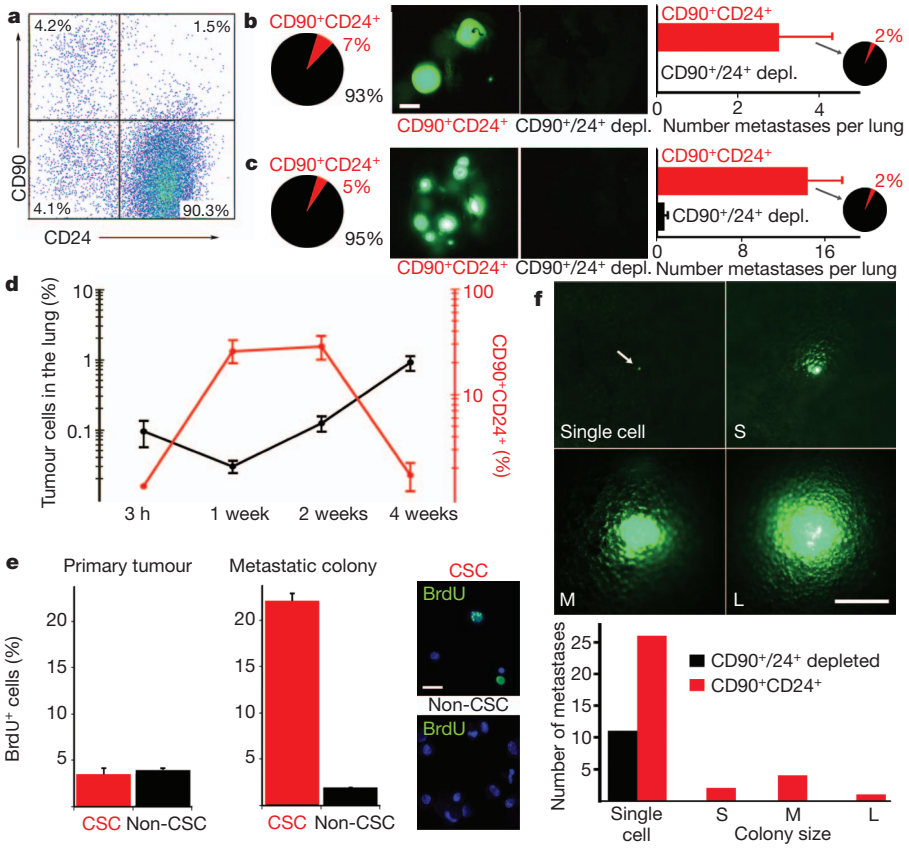


Figure 1 | Cancer stem cells initiate metastasis. (**a**) Representative density plot showing the abundance of cancer stem cells in MMTV-PyMT breast tumours defined as $CD24^+CD90^+$ after gating for viable (7-AAD^-) and lin^- ($CD31^-CD45^-TER119^-$) cells (not shown). **b**, **c**, CSCs are the only cells to form pulmonary metastases on tail vein injection. $CD24^+CD90^+$ CSCs and non-CSC populations from GFP^+ tumour cells freshly isolated from primary tumours (**b**; 10^4 cells injected each) and metastases (**c**; 10^5 cells injected each) were separately injected in recipient mice. The frequency of CSCs (red; $P < 0.01$ (**b**), $P < 0.05$ (**c**)) is maintained at the metastatic site (shown as pie charts; $n = 6$ each; errors, s.e.m.). Scale bar, 1 mm. **d**, Time course experiments show selective proliferation of CSCs during metastatic colonization. GFP^+ tumour cells (10^6) were intravenously injected into recipient mice. At the indicated time points, the total proportion of GFP^+ tumour cells in the lung (black line) and the relative amount of the CSC population (red line) were evaluated by analysis using fluorescence-activated cell sorting (FACS). Note the transient increase of CSCs after one week ($n = 4$ per time point; errors, s.e.m.). **e**, **f**, CSCs are the only cells able to initiate growth in a short-term *in vivo* colonization assay. **e**, One week after tail vein injection of 10^6 GFP^+ tumour cells, mice were injected with BrdU. After 2 h, CSCs and non-CSCs were isolated by FACS and the frequency of proliferating cells was evaluated by BrdU staining on cytospins ($n = 12$, $P < 0.05$; errors, s.d.; representative example is shown on the right). Note the increase of proliferation in the CSC population during the early phase of lung colonization; by contrast, proliferation does not significantly differ between the two populations in the primary tumour. **f**, GFP^+ CSCs (2×10^5) or GFP^+ non-CSCs (4×10^5) were tail vein injected and analysed after two weeks. $CD90^+$/$CD24^+$-depleted non-CSCs remained as single cells whereas CSCs were able to initiate growth and form metastases of different sizes (S, 30–300 cells; M, 300–3,000 cells; L, >3,000 cells). The upper panel shows a representative example of the different metastasis colonies and the results are cumulatively quantified in the bottom panel ($n = 6$). Scale bars, 20 μm (**e**) and 250 μm (**f**).

by SMAD, NF-κB and LEF/TCF. Primary lung fibroblasts upregulate POSTN in response to TGF-β3 and TGF-β2, but are not responsive to BMP4, Wnt3A, Wnt5A or TNF-α (Fig. 3e, Supplementary Fig. 19 and data not shown). Co-culture experiments revealed that tumour cells are sufficient to trigger POSTN production (Supplementary Fig. 20). Notably, CSCs and non-CSCs both produce high levels of TGF-β3 (Supplementary Fig. 21), and blocking the action of TGF-β3 by expression of a secreted decoy receptor[22] (TGFβR2ΔTM; see Methods) in tumour cells blocks POSTN expression and prevents metastasis formation (Fig. 3f and Supplementary Fig. 22) in line with earlier results[23]. Together, these experiments demonstrate that infiltrating tumour cells need to educate the host stroma of the target organ to support metastasis initiation.

Even small metastatic colonies are strongly diminished in POSTN-deficient animals (Supplementary Fig. 23). Having shown that metastatic colonization depends on cancer stem cells, we went on to assess a potential role of POSTN in stem cell maintenance. Growth under conditions of ultralow attachment has been used to study CSCs *in vitro*[24]. Surprisingly, establishment of such cancer stem cell cultures is not possible using POSTN-deficient tumours (Fig. 3g–j; $n = 16$; compare with Supplementary Fig. 24, which shows unaltered cell survival in standard two-dimensional culture). In wild-type tumour spheres, POSTN is expressed by tumour-derived, stromal fibroblasts (Supplementary Fig. 25). Adding periostin protein to the mutant tumour cells rescues sphere formation (Fig. 3j). Conversely, a blocking antibody targeting POSTN prevents maintenance of wild-type cancer stem cells (Supplementary Fig. 30), but affects neither cellular survival nor growth *per se* (Supplementary Figs 26 and 27). Furthermore, CSCs fail to proliferate when co-cultured with POSTN-mutant, pulmonary fibroblasts (Fig. 3k). In lung metastases, we observe a preferential localization of $CD90^+$ CSCs adjacent to stromal niches (Fig. 3l), whereas lung metastases in POSTN-deficient animals show a reduction in the size of the CSC population (Fig. 3m). Thus, POSTN is an essential niche factor that supports CSC growth during metastatic colonization.

To gain insight into how POSTN is related to the mechanisms that control stem cell maintenance, we characterized the interactome of POSTN by tandem affinity purification (TAP)-tag enrichment and

©2011 Macmillan Publishers Limited. All rights reserved

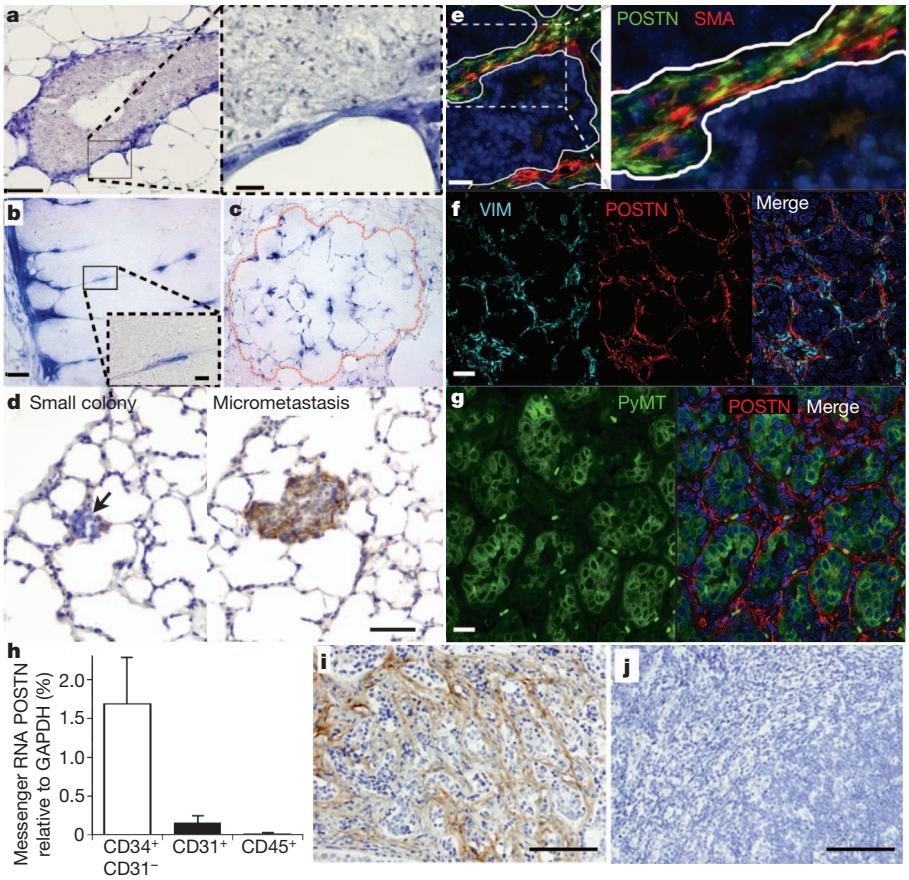


Figure 2 | POSTN is a stromal niche component that is induced on metastasis formation. **a**–**c**, Stromal POSTN expression as detected by RNA *in situ* hybridization (POSTN-expressing cells in blue) around terminal end buds, which are enriched for mammary gland stem cells during puberty (**a**), tumour cells at the primary site (**b**) and in pulmonary metastases (**c**; outlined in red). **d**, Immunohistochemistry of POSTN expression in pulmonary metastases at different stages. **e**–**g**, Immunofluorescence analysis of POSTN and fibroblast markers in pulmonary metastases. POSTN is not expressed by PyMT tumour cells (**g**), but is from αSMA$^+$ (**e**) and VIM$^+$ (**f**) fibroblasts. The white outline in **e** defines the tumour–stroma border. **h**, Quantitative PCR with reverse transcription detects POSTN expression in CD34$^+$CD31$^-$ pulmonary fibroblasts, in CD31$^+$ endothelial cells but not in CD45$^+$ immune cells that were isolated by FACS from lungs of mice with macrometastases ($n = 3$; errors, s.d.). **i**, **j**, Immunohistochemistry of POSTN expression in auxiliary lymph nodes of human breast cancer patients with metastasis-free N0 disease (**j**) or metastatic disease (**i**). A total of 17 of 23 lymph node metastases were positive for POSTN expression. Scale bars: 20 µm (**a**) and 2.5 µm (**a**, inset); 50 µm (**b**; same scale in **c**) and 10 µm (**b**, inset); 50 µm (**d**); 20 µm (**e**–**g**); 200 µm (**i**, **j**).

tandem mass spectrometry (Supplementary Methods), and this pointed to binding of POSTN to Wnt ligands[25]. Using TAP-tag pull-downs, we confirmed interaction between POSTN and Wnt1 and Wnt3A, but found that a known inhibitory ligand of the pathway, DKK1, does not associate with POSTN (Fig. 4a, b and Supplementary Fig. 28). The interaction between POSTN and Wnt ligands boosts Wnt signalling activity as measured by LRP6 phosphorylation and reporter assays using Wnt1, Wnt2 and Wnt6 (Fig. 4c, d). The Wnt pathway is known to control stem cell maintenance in a variety of tissues, including mammary gland[26] and tumours[8,27,28]. Indeed, addition of Wnt3A can rescue *in vitro* stem cell expansion in the absence of POSTN (Supplementary Figs 30 and 31). Furthermore, we find Wnt signalling activity to be concentrated in the CSC population *in vivo* as analysed by FACS (Fig. 4e, f). Here, we use a lentiviral reporter that allows Wnt-inducible GFP expression and constitutive labelling of individual reporter-containing cells by human CD2 (Supplementary Fig. 29). Wnt signalling activity in metastases is abrogated in the absence of POSTN (Fig. 4g; see Supplementary Fig. 29 for Wnt activity in tumour spheres). Similarly, we detect higher levels of the general Wnt target gene *Axin2* in metastases in wild-type hosts than in mutant hosts (Fig. 4h, i). Remarkably, metastasis of Wnt-driven breast cancer proves to be independent of stromal POSTN (Fig. 4j). Thus, POSTN acts as a niche component that can promote stem cell maintenance and metastatic colonization by augmenting Wnt signalling (Supplementary Fig. 1).

Thus, instrumental factors that contribute to the inefficiency of the metastatic process are the low abundance of CSCs, which are required to initiate growth at the secondary site, and incompatibilities with 'foreign' niches, which necessitate education of the target organ to resemble the primary microenvironment. Periostin is an essential component of this CSC-supportive niche and needs to be induced in the lung stroma by infiltrating tumour cells. Together with a recent report that implicated tumour-cell-derived tenascin C in breast cancer metastasis, these examples demonstrate an essential role for single proteins of the extracellular matrix in metastatic colonization[29,30].

Surprisingly, POSTN deficiency affects neither normal mammary gland development nor tumour formation or the relative size of the CSC population at the primary site (Supplementary Fig. 32). This may be due to a greater complexity and redundancy in the primary niche, such that the loss of a single factor can be tolerated. By contrast, loss of only one important factor, such as POSTN, prevents metastatic colonization in the presumably less supportive and less complex secondary site. Accordingly, the early phase of metastasis can be anticipated to be particularly sensitive to therapeutic intervention because the dependence of cancer cells on niche signals is probably highest in that phase. Targeting this metastatic niche promises to be less sensitive to rapid genetic changes in cancer cells and may not only prevent metastatic colonization but may also interfere with the survival of disseminated, dormant cancer cells.

©2011 Macmillan Publishers Limited. All rights reserved

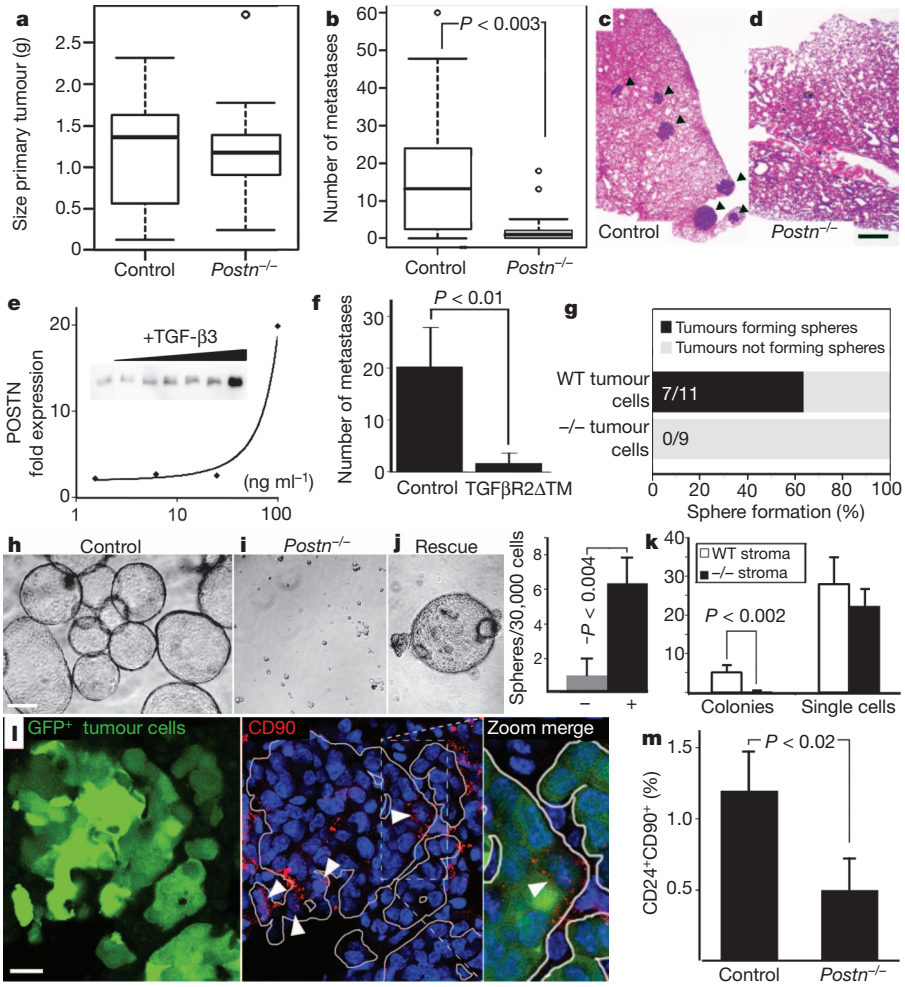


Figure 3 | POSTN is required for metastatic colonization by affecting CSC maintenance. **a**, **b**, Primary tumour size (**a**) and number of spontaneously formed macrometastases (**b**) in control and *Postn*$^{-/-}$ mice ($n = 37$). **c**, **d**, HE section of spontaneous, pulmonary metastases in control (**c**) and POSTN-mutant (**d**) animals. **e**–**f**, TGF-β3 induces POSTN expression in lung stromal cells. **e**, TGF-β3 was added to lung stromal cells to test its ability to induce POSTN expression. RNA expression by quantitative PCR and protein expression by western blot analysis (inset) were measured at the indicated concentrations. **f**, Blocking of TGF-β3 abrogates metastasis formation. Cancer cells were infected either with a control virus or with a virus expressing secreted, dominant negative TGFβR2ΔTM and separately intravenously injected into recipient mice. Metastasis formation was evaluated 6 weeks after injection ($n = 12$; errors, s.d.). **g**, Frequency of mammosphere formation from primary tumours. WT, wild type. **h**, Representative picture of control tumour mammospheres. **i**, Loss of CSCs in secondary mammosphere assays from *Postn*$^{-/-}$ tumours. **j**, Sphere formation by *Postn*$^{-/-}$ tumour cells can be rescued by addition of 50 ng ml^{-1} of periostin to primary cultures. Control tumour cells produce 13.3 ± 2.5 spheres per 30,000 cells. Quantification is shown in the histogram on the right ($n = 4$; errors, s.d.). **k**, Only stromal cells isolated from wild-type lungs are able to support CSC colony formation *in vitro*. CSCs were sorted by FACS and seeded on lung stromal cells isolated either from wild-type or *Postn*$^{-/-}$ mice. Colony formation was evaluated after two weeks ($n = 3$; errors, s.d.). **l**, Immunofluorescence analysis shows that GFP$^+$CD90$^+$ CSCs (white arrow heads) are concentrated in areas with immediate contact to the stroma. **m**, Reduced frequency of CD24$^+$CD90$^+$ CSCs in rare pulmonary metastases in POSTN-deficient mice ($n = 12$; errors, s.d.). Scale bars: 500 μm (**c**, **d**); 50 μm (**h**–**j**); 5 μm (**l**).

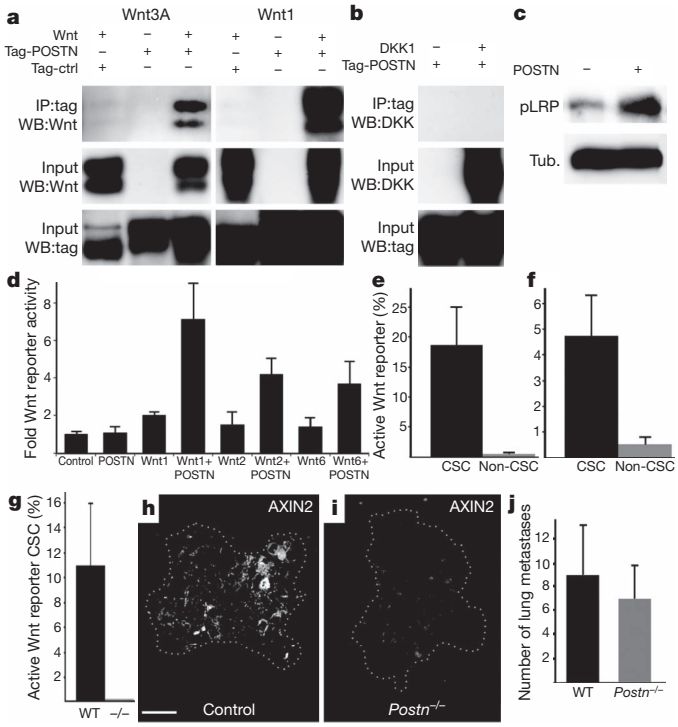


Figure 4 | POSTN promotes stem cell maintenance and metastasis formation by augmenting Wnt signalling. **a**, Interaction of TAP-tagged POSTN and Wnt3A and Wnt1 as determined by pull down and western blotting. **b**, Interaction of TAP-tagged POSTN and DKK1 as determined by pull down and western blotting. Note the absence of interaction between POSTN and DKK1. **c**, LRP6 phosphorylation (pLRP) increases on POSTN expression. **d**, Wnt reporter activity in response to Wnt1, Wnt2 or Wnt6 is stimulated by POSTN ($n = 3$; errors, s.d.). **e**, **f**, *In vivo* Wnt reporter assay. Tumour cells were infected with a lentivirus containing both an artificial Wnt-responsive promoter driving GFP expression and a constitutive promoter expressing human CD2 (Supplementary Fig. 29). Cells were analysed by FACS considering only human-CD2$^+$, reporter-containing cells. CD90$^+$CD24$^+$ CSCs and non-CSCs were scored for Wnt reporter activity as GFP$^+$. The signal concentrates in CSCs in both primary tumours (**e**) and pulmonary metastases (**f**) ($n = 3$; errors, s.e.m.). **g**, CD90$^+$CD24$^+$ wild-type CSCs containing a Wnt reporter (human CD2$^+$) show signalling activity (GFP$^+$) only when seeded in wild-type as opposed to *Postn*$^{-/-}$ lungs ($n = 3$; errors, s.e.m.). **h**, **i**, Immunofluorescence staining of AXIN2 in control (**h**) and POSTN-deficient (**i**) lung metastases (outlined in white). Note the absence of staining in POSTN$^-$ lung metastases. Scale bar, 10 μm. **j**, MMTV-Wnt1 tumour cells can form pulmonary metastases in the absence of POSTN expression in the stroma. Tumour cells were orthotopically transplanted into mammary glands of wild-type or POSTN-deficient recipients and analysed after ten weeks for spontaneous lung metastasis ($n = 5$ each; errors, s.d.).

©2011 Macmillan Publishers Limited. All rights reserved

METHODS SUMMARY

Details on the mouse models used and the generation of the *Postn* gene ablation can be found in Supplementary Information. Tumours derived from MMTV-PyMT mice were collected and cells were isolated and directly stained for FACS analysis or sorting using antibodies against CD31, CD45, TER119 (LY76), CD90, CD24 and CD49f (ITGA6). FACS-isolated tumour cells were directly transplanted into recipient mice.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 July 2010; accepted 1 November 2011.
Published online 7 December 2011.

1. Chambers, A. F., Groom, A. C. & MacDonald, I. C. Dissemination and growth of cancer cells in metastatic sites. *Nature Rev. Cancer* **2,** 563–572 (2002).
2. Hüsemann, Y. *et al.* Systemic spread is an early step in breast cancer. *Cancer Cell* **13,** 58–68 (2008).
3. Kouros-Mehr, H. *et al.* GATA-3 links tumor differentiation and dissemination in a luminal breast cancer model. *Cancer Cell* **13,** 141–152 (2008).
4. Nguyen, D. X., Bos, P. D. & Massague, J. Metastasis: from dissemination to organ-specific colonization. *Nature Rev. Cancer* **9,** 274–284 (2009).
5. Hess, K. R. *et al.* Metastatic patterns in adenocarcinoma. *Cancer* **106,** 1624–1633 (2006).
6. Lin, E. Y. *et al.* Progression to malignancy in the polyoma middle T oncoprotein mouse breast cancer model provides a reliable model for human diseases. *Am. J. Pathol.* **163,** 2113–2126 (2003).
7. Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. *Proc. Natl Acad. Sci. USA* **100,** 3983–3988 (2003).
8. Reya, T. & Clevers, H. Wnt signalling in stem cells and cancer. *Nature* **434,** 843–850 (2005).
9. Cho, R. W. *et al.* Isolation and molecular characterization of cancer stem cells in MMTV-Wnt-1 murine breast tumors. *Stem Cells* **26,** 364–371 (2008).
10. Shackleton, M. *et al.* Generation of a functional mammary gland from a single stem cell. *Nature* **439,** 84–88 (2006).
11. Stingl, J. *et al.* Purification and unique properties of mammary epithelial stem cells. *Nature* **439,** 993–997 (2006).
12. Liu, J. C., Deng, T., Lehal, R. S., Kim, J. & Zacksenhaus, E. Identification of tumorsphere- and tumor-initiating cells in HER2/Neu-induced mammary tumors. *Cancer Res.* **67,** 8671–8681 (2007).
13. Zhang, M. *et al.* Identification of tumor-initiating cells in a p53-null mouse model of breast cancer. *Cancer Res.* **68,** 4674–4682 (2008).
14. Podsypanina, K. *et al.* Seeding and propagation of untransformed mouse mammary cells in the lung. *Science* **321,** 1841–1844 (2008).
15. Vermeulen, L. *et al.* Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. *Nature Cell Biol.* **12,** 468–476 (2010).
16. Visvader, J. E. & Lindeman, G. J. Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. *Nature Rev. Cancer* **8,** 755–768 (2008).
17. Psaila, B. & Lyden, D. The metastatic niche: adapting the foreign soil. *Nature Rev. Cancer* **9,** 285–293 (2009).
18. Takeshita, S., Kikuno, R., Tezuka, K. & Amann, E. Osteoblast-specific factor 2: cloning of a putative bone adhesion protein with homology with the insect protein fasciclin I. *Biochem. J.* **294,** 271–278 (1993).
19. Shimazaki, M. & Kudo, A. Impaired capsule formation of tumors in periostin-null mice. *Biochem. Biophys. Res. Commun.* **367,** 736–742 (2008).
20. Oka, T. *et al.* Genetic manipulation of periostin expression reveals a role in cardiac hypertrophy and ventricular remodeling. *Circ. Res.* **101,** 313–321 (2007).
21. Rios, H. *et al.* Periostin null mice exhibit dwarfism, incisor enamel defects, and an early-onset periodontal disease-like phenotype. *Mol. Cell. Biol.* **25,** 11131–11144 (2005).
22. Zhao, W. *et al.* Suppression of *in vivo* tumorigenicity of rat hepatoma cell line KDH-8 cells by soluble TGF-beta receptor type II. *Cancer Immunol. Immunother.* **51,** 381–388 (2002).
23. Muraoka, R. S. *et al.* Blockade of TGF-beta inhibits mammary tumor cell viability, migration, and metastases. *J. Clin. Invest.* **109,** 1551–1559 (2002).
24. Pece, S. *et al.* Biological and molecular heterogeneity of breast cancers correlates with their cancer stem cell content. *Cell* **140,** 62–73 (2010).
25. Milovanovic, T. *et al.* Expression of Wnt genes and frizzled 1 and 2 receptors in normal breast epithelium and infiltrating breast carcinoma. *Int. J. Oncol.* **25,** 1337–1342 (2004).
26. Zeng, Y. A. & Nusse, R. Wnt proteins are self-renewal factors for mammary stem cells and promote their long-term expansion in culture. *Cell Stem Cell* **6,** 568–577 (2010).
27. Malanchi, I. *et al.* Cutaneous cancer stem cell maintenance is dependent on β-catenin signaling. *Nature* **452,** 650–653 (2008).
28. Barker, N. *et al.* Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457,** 608–611 (2009).
29. Wels, J., Kaplan, R. N., Rafii, S. & Lyden, D. Migratory neighbors and distant invaders: tumor-associated niche cells. *Genes Dev.* **22,** 559–574 (2008).
30. Oskarsson, T. *et al.* Breast cancer cells produce tenascin C as a metastatic niche component to colonize the lungs. *Nature Med.* **17,** 867–874 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to U. Koch for advice on FACS, to M. Moniatte for advice on mass spectrometry and to S. Leuba for technical assistance with histology. I.M., E.S., A.S.-M. and J.H. were supported in part by the Swiss League against Cancer, the SNF, the NCCR in Molecular Oncology and the Anna Fuller Fund. J.H. holds the EPFL chair for Signal Transduction in Oncogenesis sponsored by Debiopharm.

Author Contributions I.M., A.S.-M. and J.H. designed and performed most of the experiments, analysed data and prepared the manuscript; E.S. and H.P. performed experiments; H.-A.L. performed experiments and analysed data; J.-F.D. provided clinical samples; and J.H. designed and supervised the study.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.H. (joerg.huelsken@epfl.ch).

©2011 Macmillan Publishers Limited. All rights reserved

METHODS

Mouse work. The design of the POSTN targeting vector is shown in Supplementary Fig. 10. E14.1 129/Ola ES cells were used for targeting and were injected into C57Bl6 blastocysts. The POSTN mutant allele was then backcrossed onto the FVB background for at least eight generations. The MMTV-PyMT[31], MMTV-Wnt1[32] and the ACT–GFP[33] mouse strains have been described previously. All used strains were backcrossed onto FVB for ten generations except for the Nude allele, which was used on a NMRI background. For tumour cell transplantations, *Nude* or *Rag1*$^{-/-}$ mice were used when tumour cells expressed GFP or had been modified by lentiviral vectors. In general, we did not observe major differences in these different strains with respect to cancer stem cell phenotype, tumour take rate or metastasis number, except for an overall faster growth in *Rag1*$^{-/-}$ versus *Nude* versus wild type. For tail vein injections, tumour cells were resuspended in 100 µl of PBS. For orthotopic transplantations, tumour cells were resuspended in 50 µl growth-factor-reduced matrigel (BD Biosciences) and transplanted into a small pocket within the fourth mammary fat pad. All animal procedures were performed in accordance with the Swiss legislation on animal experimentation.

Antibodies. We used TER119, CD24, CD29, CD34, CD45 and CD90.1 (eBioscience); CD31 and GFP (Invitrogen); Cytokeratin 8, POSTN, AXIN2 and Wnt1 (Abcam); POSTN (Adipogen); Cytokeratin 14 (Covance); GATA-3 and PyMT (Santa Cruz Biotechnology); αSMA (Sigma); VIM (Lifespan Biosciences); and BrdU (BD Pharmingen and ImmunologicalsDirect). The antibody against human CD2 was a kind gift by Professor Werner Held (Ludwig Center for Cancer Research, Lausanne).

Tumour and metastasis evaluation. For the quantification of spontaneous micrometastasis, lungs were completely sectioned and HE stained, and the size and the number of metastatic nodules were analysed by microscopy. Quantification of lung metastases derived from GFP^{+} tumour cells was performed by fluorescent microscopy (Leica M205 FA stereomicroscope) or by FACS analysis of cell preparations from total lungs. All statistical evaluation used a homoscedastic Student's *t*-test with a two-tailed distribution.

For expression of a secreted TGFβ decoy receptor (TGFβR2ΔTM), tumour cells were transduced with lentiviral vectors expressing the extracellular domain (amino acids 1–185) of murine TGFβR2 (NM_009371.3) under the control of a murine phosphoglycerate kinase (PGK) promoter, and metastases were scored as described above following tail vein injection of transduced tumour cells.

Tumour cell isolation. Tumours or lungs containing pulmonary metastases were minced with a razor blade and digested with a mixture of DNase and Liberase (Roche Diagnostics). On enzymatic digestion, samples were passed through a 100-µm filter and washed once in growth medium (DMEM/F12 with 2% FBS, 20 ng ml^{-1} EGF, 10 µg ml^{-1} insulin; Invitrogen) and twice in PBS. Cells were then directly used for FACS staining and analysed in a Cyan ADP flow cytometer (Beckman Coulter) or sorted with a custom-built FACSAria II (Beckton Dickinson). For tail vein injection, cells were plated overnight on collagen. In the experiments evaluating the metastasis initiation potential of CSCs versus non-CSCs, we were able to prepare and inject higher cell numbers from metastases than from primary tumours, where larger necrotic areas limit the obtainable cell amounts. In both settings, only CSCs were able to give rise to pulmonary metastases, but this experimental setting is unsuitable for a direct comparison of the metastatic potential of CSCs isolated from primary tumours with that of those isolated from metastases. In the extravasation experiments, we performed several pilot experiments perfusing the lungs before dissection to distinguish cells trapped inside capillaries from those that successfully extravasated. Perfusion was done with Ringer's buffer through the right ventricle for 5–10 min.

Tumour mammosphere cultures. Sphere cultures were established from total tumour cell preparations. After dissociation, tumour cells were plated on collagen overnight, trypsinized and plated in 100 µl of sphere media (DMEM/F12 with B27, 20 ng ml^{-1} EGF, 20 ng ml^{-1} FGF and 4 µg ml^{-1} heparin) into 96-well, low-attachment plates (Constar) at a concentration of (1–3) $\times$ 10^{4} cells per well. For the anti-POSTN blocking antibody assay, half of the culture medium was replaced by hybridoma supernatant. Secondary mammosphere culture was performed by collecting the spheres through gentle centrifugation (130*g*) followed by enzymatic (10 min in 0.05% trypsin-EDTA at 37 °C) and mechanical dissociation using a 25G needle. Immunostainings were performed as whole mounts of PFA-fixed and 0.1% Triton-X100-permeabilized spheres and were analysed by confocal microscopy.

Identification of POSTN as a niche-expressed factor. To identify niche-expressed factors, we performed microarray RNA profiling studies where we analysed laser-capture microdissected material from normal stem cells together with their adjacent stroma. This was compared with laser-capture microdissected material isolated from differentiated cells together with their adjacent stroma from the same tissue (data not shown). Skin and intestine contain stem cells in morphologically distinct structures (the bulge in the skin, the crypt in the intestine). On the basis of these profiles (accessible through GEO GSE31730 and 8818), we evaluated potential niche localization of candidate genes by RNA *in situ* hybridization and immunofluorescence analysis (Fig. 2a, Supplementary Fig. 7 and data not shown). Moreover, we profiled the candidate genes by quantitative PCR with reverse transcription and performed RNA *in situ* hybridization in other tissues, in tumours and in metastases (data not shown). This identified POSTN as a stromal marker in primary murine breast cancer and its metastases (Fig. 2).

Immunodetection and *in situ* hybridization. Immunostaining was performed on 7-µm OCT- or gelatin-embedded (to maintain GFP signals) frozen sections permeabilized with 0.1% Triton-X100 or on 5-µm paraffin sections using antigen retrieval for 20 min in boiling, 10 mM citric acid, pH 6.0. For BrdU immunostaining, an additional incubation period of 10 min in 2 N HCl was performed. For immunohistochemistry, endogenous peroxidases were blocked by incubation with 0.6% H_2O_2 in methanol. After incubation with the indicated antibodies, secondary fluorescently labelled antibodies Alexa Fluor 488, 567 and 647 (Molecular Probes, Invitrogen) or HRP-conjugated secondary antibodies were used to reveal the primary antibodies. For immunohistochemistry, Dako Envision^{+} was used together with diaminobenzidine tablets (Sigma) to detect HRP. Fluorescent images were made with an automated upright microscope system (DM5500, Leica) or an LSM700 upright or inverted confocal microscope (Zeiss). Light images were made with an AX70 widefield microscope (Olympus). Figures were generated with PHOTOSHOP (Adobe Systems) and CYTOSKETCH (CytoCode).

In situ hybridizations were performed on paraffin sections permeabilized by proteinase K digestion using DIG-labelled (Roche), antisense transcripts of the mouse POSTN complementary DNA. After hybridization, section were washed in 50% formamide, ×5 SSC (0.75 M NaCl, 85 mM sodium citrate, pH 7.0) and 0.1% Tween at 70 °C, and non-specifically bound probe was digested by RnaseA treatment, followed by extensive washing in 50% formamide, ×2 SSC and 0.1% Tween at 65 °C. The DIG label was detected by an anti-DIG Fab (Roche) coupled to alkaline phosphatase using FAST NBT/BCIP (Sigma).

Western blot. Protein was extracted with complete RIPA buffer, separated by electrophoresis, transferred to PVDF membranes, blocked with 5% dry milk or BSA and incubated overnight with primary antibodies. Immunoreactive bands were visualized using HRP-conjugated secondary antibodies (Promega) and the detection reagent SuperSignal West Cheminoluminiscent Substrate (Thermo Scientific).

Real-time RT–PCR. RNA was prepared using a miniRNA or microRNA kit (Qiagen) from FACS-isolated primary cells. Complementary DNAs were generated using oligo-T priming and quantitative PCR was performed in a LightCycler (Roche) or a StepOnePlus thermocycler (Applied Biosystems) using the Power SYBR green PCR Master Mix (Applied Biosystems) and the primers aatgctgccctggc tatatg and gcatgacccttttccttcaa (POSTN); agcgcagacacaacccatag and gggttctgccca catagtacaag (TGF-β3); and caagctcatttcctggtatgacaat and gttgggatagggcctctcttg (GAPDH).

Periostin monoclonal antibody production. In collaboration with Adipogen, we injected *Postn*$^{-/-}$ mice with purified human POSTN protein. Hybridomas were generated and tested for antibody production by ELISA. Antibodies were selected for their ability to recognize POSTN and for their ability to block tumour mammosphere formation *in vitro*.

Periostin interactome and pull-downs. Freshly isolated tumour cells were grown under mammosphere conditions in the presence of purified, polyHis-tagged and streptavidin binding peptide (SBP)-tagged POSTN protein (Supplementary Fig. 28) for 72 h. Then proteins were crosslinked by 1 mM DTSSP (Pierce) for 10 min and solubilized by DDM and SDS (Anatrace). Protein purification was achieved by two successive affinity purification steps using the SBP and the polyHis tags and elution by biotin and imidazole, respectively. After elution, the crosslink was released by DTT treatment, the resulting protein mixture was trypsin-digested and peptides were identified by one-dimensional nano-liquid chromatography tandem mass spectrometry. Among the POSTN-attached proteins, we found Wnt6 to be a potential interactor. Because these experiments involved crosslinkers, we consider these results to be preliminary unless confirmed by independent experiments. For Wnt6, an interaction with POSTN could not be validated owing to a lack of Wnt6 antibodies and a suitable expression system for Wnt6. Nevertheless, this preliminary finding prompted us to test directly for interactions of POSTN with other Wnt ligands where the necessary tools are available. For these pull-downs, cDNAs encoding Wnt1, HA-tagged Wnt3A, DKK1 and SBP/His-tagged POSTN (Supplementary Fig. 28) or SBP/His-tagged mouse immunoglobulin-G heavy chain (as a control) were transfected separately into 293T cells. Cells secreting POSTN or immunoglobulin-G and cells producing Wnt or DKK were mixed after one day and collected for pull-downs after three additional days. Without the use of crosslinkers, POSTN-interacting proteins were enriched by the SBP/His tag as described above using NP40 as detergent. Western blot was performed for identification of the pulled-down proteins.

©2011 Macmillan Publishers Limited. All rights reserved

Wnt reporter activity. HEK293T cells were transduced with a lentiviral vector containing a LEF/TCF-responsive luciferase reporter consisting of a minimal promoter and 20 copies of the LEF/TCF consensus site or a control construct in which the LEF/TCF sites had been mutated[34]. These stable Wnt reporter cell lines were then transiently transfected with a POSTN expression vector before these cells were mixed with transiently transfected HEK293T or COS7 cells (the latter for Wnt6 expression) producing Wnt1, Wnt2 or Wnt6, respectively. The media on these co-cultures of Wnt reporter and Wnt producer cells were replaced after 24 h to remove soluble Wnt ligands. After 48 h of co-culture, luciferase activity in cell extracts was measured with a kit from Promega.

For *in vivo* reporter assays, we generated a lentiviral vector containing the same transcription factor sites and minimal promoter as described above driving GFP expression. In addition, this vector contained a ubiquitous PGK promoter driving expression of a truncated human CD2 cDNA that allows identification of reporter-containing cells by FACS (Supplementary Fig. 29). Lentiviruses were generated in 293T cells using third-generation lentivirus packaging vectors, and virus particles were concentrated by ultracentrifugation. Freshly isolated tumour cells were infected with lentiviruses overnight, enriched by anti-CD2 MACS (Miltenyi) for cells carrying the reporter construct and orthotopically transplanted for tumour formation and spontaneous metastasis. Wnt signalling activity in cells of primary tumours or spontaneous pulmonary metastases was quantified by FACS analysis by measuring GFP^+ cells after gating for human-$CD2^+$ cells. Similarly, human-$CD2^+$ tumour cells were used for tumour mammosphere cultures to measure Wnt pathway activity in the presence or absence of the POSTN blocking antibody.

Isolation of lung stromal cells. Lungs were minced with a razor blade and digested with a mixture of DNase and Liberase (Roche Diagnostics). On enzymatic digestion, samples were passed through a 100-μm filter and washed once in growth medium (DMEM/F12 with 2% FBS, 20 ng ml^{-1} EGF and 10 μ ml^{-1} insulin; Invitrogen) and twice in PBS. Cells were plated onto plates pretreated with collagen and allowed to attach for three days. The medium was then changed to serum-free medium and tumour cells or soluble factors were added to test POSTN induction. POSTN expression was evaluated by western blot, PCR with reverse transcription or immunofluorescence analysis.

31. Guy, C. T., Cardiff, R. D. & Muller, W. J. Induction of mammary tumours by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. *Mol. Cell. Biol.* **12,** 954–961 (1992).
32. Tsukamoto, A. S., Grosschedl, R., Guzman, R. C., Parslow, T. & Varmus, H. E. Expression of the int-1 gene in transgenic mice is associated with mammary gland hyperplasia and adenocarcinomas in male and female mice. *Cell* **55,** 619–625 (1988).
33. Okabe, M., Ikawa, M., Kominami, K., Nakanishi, T. & Nishimune, Y. 'Green mice' as a source of ubiquitous green cells. *FEBS Lett.* **407,** 313–319 (1997).
34. Jeannet, G. *et al.* Long-term, multilineage hematopoiesis occurs in the combined absence of β-catenin and γ-catenin. *Blood* **111,** 142–149 (2008).

©2011 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature10709

Reversal of cocaine-evoked synaptic potentiation resets drug-induced adaptive behaviour

Vincent Pascoli[1], Marc Turiault[1] & Christian Lüscher[1,2]

Drug-evoked synaptic plasticity is observed at many synapses and may underlie behavioural adaptations in addiction[1]. Mechanistic investigations start with the identification of the molecular drug targets. Cocaine, for example, exerts its reinforcing[2] and early neuroadaptive effects[3] by inhibiting the dopamine transporter, thus causing a strong increase in mesolimbic dopamine. Among the many signalling pathways subsequently engaged, phosphorylation of the extracellular signal-regulated kinase (ERK) in the nucleus accumbens[4] is of particular interest because it has been implicated in NMDA-receptor and type 1 dopamine (D1)-receptor-dependent synaptic potentiation[5] as well as in several behavioural adaptations[6–8]. A causal link between drug-evoked plasticity at identified synapses and behavioural adaptations, however, is missing, and the benefits of restoring baseline transmission have yet to be demonstrated. Here we find that cocaine potentiates excitatory transmission in D1-receptor-expressing medium-sized spiny neurons (D1R-MSNs) in mice via ERK signalling with a time course that parallels locomotor sensitization. Depotentiation of cortical nucleus accumbens inputs by optogenetic stimulation *in vivo* efficiently restored normal transmission and abolished cocaine-induced locomotor sensitization. These findings establish synaptic potentiation selectively in D1R-MSNs as a mechanism underlying a core component of addiction, probably by creating an imbalance between distinct populations of MSNs in the nucleus accumbens. Our data also provide proof of principle that reversal of cocaine-evoked synaptic plasticity can treat behavioural alterations caused by addictive drugs and may inspire novel therapeutic approaches involving deep brain stimulation or transcranial magnetic stimulation.

We first tested whether cocaine treatment interfered with activity-dependent long-term potentiation (LTP) in the nucleus accumbens. When excitatory afferents onto MSNs were challenged with a high-frequency stimulation (HFS) train, LTP of the excitatory postsynaptic currents (EPSCs) was observed (Fig. 1a). The magnitude of the LTP was halved in brain slices from mice that had received a single injection of cocaine 7 days before the recording. If the cocaine was injected a month before measuring synaptic plasticity *ex vivo*, the difference was no longer present (Fig. 1b). MSNs of the nucleus accumbens fall into two classes of about equal proportions defined by the type of dopamine receptor expressed, with a small fraction of neurons (6–17%) that express both receptors[9]. A possible explanation for the partial change in LTP magnitude is therefore that cocaine exposure abolishes plasticity selectively in one class. To test this, we attempted LTP induction *ex vivo* after cocaine exposure in bacterial artificial chromosome (BAC) transgenic mice expressing enhanced green fluorescent protein (eGFP) either in D1 receptor (D1R)- or D2R-MSNs. We identified D1R-MSNs by a crossover strategy in which we recorded from green cells in *drd1a*-eGFP mice and non-green cells in *drd2*-eGFP mice (and vice versa for D2R-MSNs). Because the two approaches to identify the cell type yielded very similar results, we pooled the data (Fig. 1c, d). The main finding of this first experiment was that HFS, which reliably induced LTP in both types of MSNs after saline injection, became inefficient after cocaine treatment in D1R-MSNs (Supplementary Fig. 1a–d). The D2R overexpression recently reported in *drd2*-eGFP mice[10] interferes with neither the reported synaptic effects nor with the acute locomotor response to cocaine (Supplementary Fig. 2). After a cocaine injection, we were unable to induce HFS LTP regardless of whether we recorded from MSNs in the shell or the core of the nucleus accumbens (Supplementary Fig. 3), but because MSNs in the shell are the immediate targets of the dopamine neurons of the medial ventral tegmental area, which undergo the most significant changes in response to a single cocaine injection[11], we focused on nucleus accumbens shell neurons in the present study. Without distinguishing between D1R- and D2R-MSNs, several reports have already suggested that drug-evoked synaptic plasticity in the nucleus accumbens may underlie drug-related behavioural adaptations (reviewed in ref. 12). For example, when 5–7 daily cocaine injections were followed by a 10–21-day withdrawal, an overall increase of the AMPA/NMDA receptor (AMPAR/NMDAR) ratio[13] or GluA1/2 surface expression[14,15] was observed, and both observations were reversed by a challenge injection of cocaine.

The failure of HFS to induce LTP selectively in D1R-MSNs after cocaine treatment may be due either to the impairment of LTP induction or the occlusion of LTP expression. To distinguish between the two scenarios, we recorded miniature EPSCs (mEPSCs) in both cell types (Fig. 1e and Supplementary Fig. 4a) and observed a significant increase in amplitude along with a modest change in frequency of unitary events in D1R-MSNs (whereas these parameters remained unchanged in D2R-MSNs). Given that the paired pulse ratio also remained unchanged (Fig. 1f and Supplementary Fig. 4b), a post-synaptic mechanism underlying the increase of transmission and hence an occlusion scenario is the most likely explanation.

We characterized the induction criteria for HFS LTP in MSNs *in vitro* and found that it depends on NMDAR activation (Fig. 2a). Applying the MEK (MAP ERK kinase) inhibitor U0126 for the duration of the induction protocol also led to a complete block (Fig. 2b), indicating that activation of the ERK pathway is essential for this form of LTP, akin to NMDA-dependent LTP in the hippocampus and the dorsal striatum[16,17], the latter also depending on D1Rs[18].

Given that our results indicate an occlusion of LTP, we hypothesized that cocaine could drive synaptic potentiation selectively in D1R-MSNs via ERK activity. Indeed, a sharp increase in phosphorylated ERK is detected soon after cocaine exposure in D1R-MSNs, but not in D2R-MSNs[9]. In the nucleus ERK modulates gene expression, whereas in dendrites ERK is probably involved in the regulation of activity-dependent spine dynamics, synaptic glutamate receptor insertion and local dendritic protein synthesis[19]. To provide *in vivo* evidence for ERK dependence of cocaine-evoked synaptic plasticity, we treated mice with SL327, a blood–brain barrier penetrant ERK pathway inhibitor, before the saline or cocaine injection. We found that this manipulation rescued HFS LTP in D1R-MSNs one week later (Fig. 2c, d) without modification of the acute locomotor response to cocaine

[1]Department of Basic Neurosciences, Medical Faculty, University of Geneva, CH-1211 Geneva, Switzerland. [2]Clinic of Neurology, Department of Clinical Neurosciences, Geneva University Hospital, CH-1211 Geneva, Switzerland.

©2011 Macmillan Publishers Limited. All rights reserved

(Supplementary Fig. 5). A crossover control design with the two mouse lines again showed no difference between the green cells of one line and the non-green cells of the other (Supplementary Fig. 6).

Because inhibition of the ERK pathway blocks locomotor sensitization to cocaine[20,21], and we found LTP in D1R-MSNs to be dependent on ERK, we reasoned that cocaine-evoked potentiation might be a cellular correlate of the behavioural adaptation. Moreover, ERK activation is correlated both with AMPAR expression at the cell surface throughout the nucleus accumbens and with locomotor sensitization[14], but the synapses involved have not been identified, and no causal link has been established.

We then confirmed that a single injection of cocaine was sufficient to cause locomotor sensitization to a second injection of the same dose[22]. This was the case when challenged a week but not a month after the initial cocaine injection (Supplementary Fig. 7). The behavioural alteration therefore followed a time course similar to the cocaine-evoked synaptic potentiation (Fig. 1a, b). If cocaine-evoked potentiation is causally involved in locomotor sensitization, then depotentiating these synapses may reverse the behavioural alterations. To test this prediction experimentally, we injected channelrhodopsin (ChR2)-expressing adeno-associated virus (AAV) into the infralimbic cortex and implanted light guides into the nucleus accumbens to be able to selectively activate *in vivo* the terminals of this major excitatory input. Histological verification of the injection site did confirm robust expression in the infralimbic cortex, along with sparser expression in the prelimbic cortex (Fig. 3a and Supplementary Fig. 8). To validate further this approach, we recorded photocurrents from infected cortical neurons and prepared slices of the nucleus accumbens shell in which wide-field light exposure led to robust AMPAR-mediated EPSCs (Fig. 3b). We next applied light pulses at 1 Hz for 10 min (an established long-term depression (LTD) protocol to reduce synaptic transmission, see Methods). We found that this protocol strongly depressed transmission in nucleus accumbens slices from both saline and cocaine-treated mice (Fig. 3c). Interestingly, the magnitude of the depression was significantly larger in the latter, in line with an efficient depotentiation added to the LTD. As the LTD/depotentiation was NMDAR dependent and occurred without change of the paired pulse ratio, it is probably mediated by a postsynaptic expression mechanism and therefore constitutes an actual reversal of cocaine-evoked potentiation (Supplementary Fig. 9).

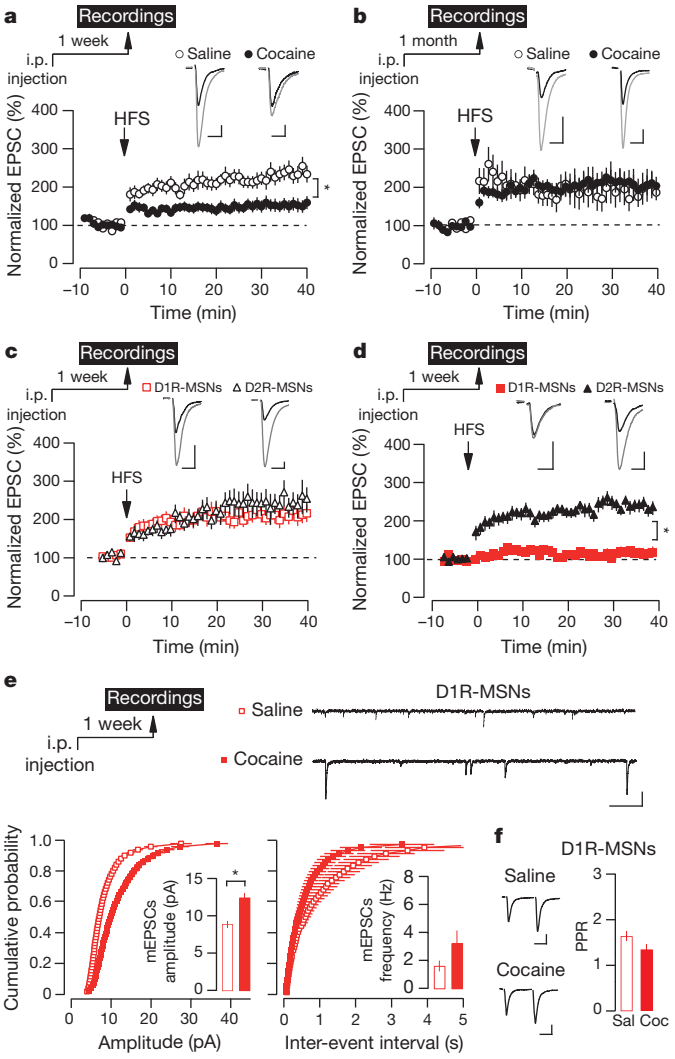


Figure 1 | Cocaine disrupts HFS-induced LTP in D1R-MSNs of the nucleus accumbens by potentiation of excitatory afferents. a–d, Graphs show normalized EPSCs as a function of time and the overlay of averaged (20 trials) traces of AMPAR EPSCs before (black line) and after (grey line) HFS. Symbols represent average of 6 trials. **a**, One week after cocaine injection, LTP was halved (224 ± 17.9% to 149 ± 14.7%, Student's *t*-test $t_{43} = 3.06$). **b**, After 1 month, no difference between treatments was detected (204 ± 26.1% to 206 ± 21.3%, $t_{21} = 0.07$). **c**, Using *drd1a*- and *drd2*-eGFP mice, HFS reliably induced LTP in D1R-MSNs and D2R-MSNs after saline injection (206 ± 16.0% and 229 ± 24.1%, respectively, $t_{27} = 0.78$). **d**, Cocaine abolished HFS LTP in D1R-MSNs but not in D2R-MSNs (115 ± 8.5% and 239 ± 12.7%, respectively, $t_{39} = 7.98$). **e**, Sample traces, cumulative probability and mean values of amplitude and frequency for mEPSCs recorded from D1R-MSNs one week after injection of saline or cocaine. Cocaine significantly increases mean amplitude of mEPSCs in D1R-MSNs ($t_{14} = 4.68$; Kolmogorov–Smirnov test: $P < 0.01$) whereas the frequency was unaffected ($t_{14} = 1.635$; Kolmogorov–Smirnov: $P > 0.05$). **f**, Paired-pulse ratio (PPR, 50 ms inter-stimulus interval) measured in D1R-MSNs was not different a week after saline (Sal) or cocaine (Coc) injection (1.7 ± 0.11 and 1.4 ± 0.1, respectively, $t_{34} = 1.73$). Scale bars: 20 ms and 20 pA. $n = 8–30$, $^*P < 0.05$. Error bars show s.e.m.

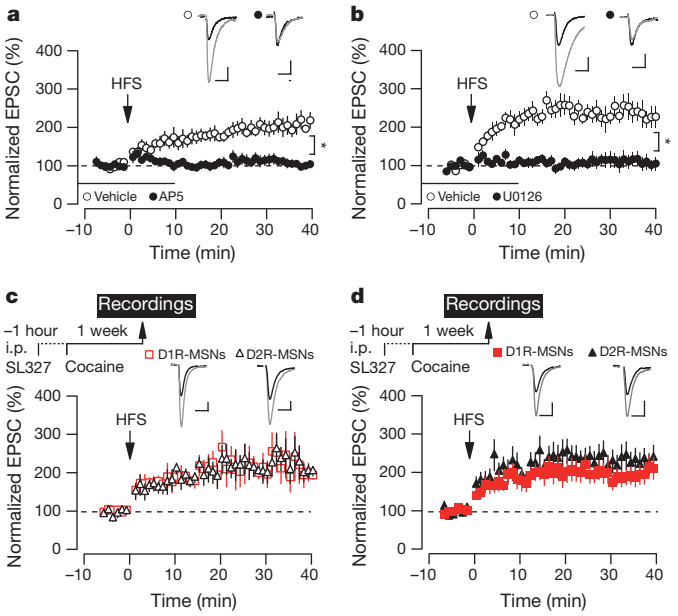


Figure 2 | HFS LTP and cocaine-evoked potentiation both depend on ERK activation. a, HFS LTP was blocked by an NMDAR antagonist (AP5, 100 μM; control, 209 ± 15% versus AP5, 93 ± 9%, $t_{14} = 6.78$). **b**, Bath application of the MEK inhibitor (U0126, 5 μM, 15 min before until 10 min after HFS protocol) also blocked LTP (control, 234 ± 17.8% versus U0126, 109 ± 16.2%, $t_{13} = 5.047$). **c**, Intraperitoneal administration of the MEK inhibitor (SL237, 40 mg kg^{-1}) in *drd1a*-eGFP mice or *drd2*-eGFP mice 1 h before saline did not modify the magnitude of HFS LTP in D1R- or D2R-MSNs when assessed one week later (224 ± 31% and 221 ± 27%, respectively, $t_{10} = 0.09$). **d**, SL237, administered 1 h before cocaine restored HFS LTP in D1R-MSNs (200 ± 22% versus 232 ± 27% for D2R-MSNs, $t_{22} = 0.93$). Scale bars: 10 ms and 20 pA. $n = 6–13$, $^*P < 0.05$. Error bars show s.e.m.

©2011 Macmillan Publishers Limited. All rights reserved

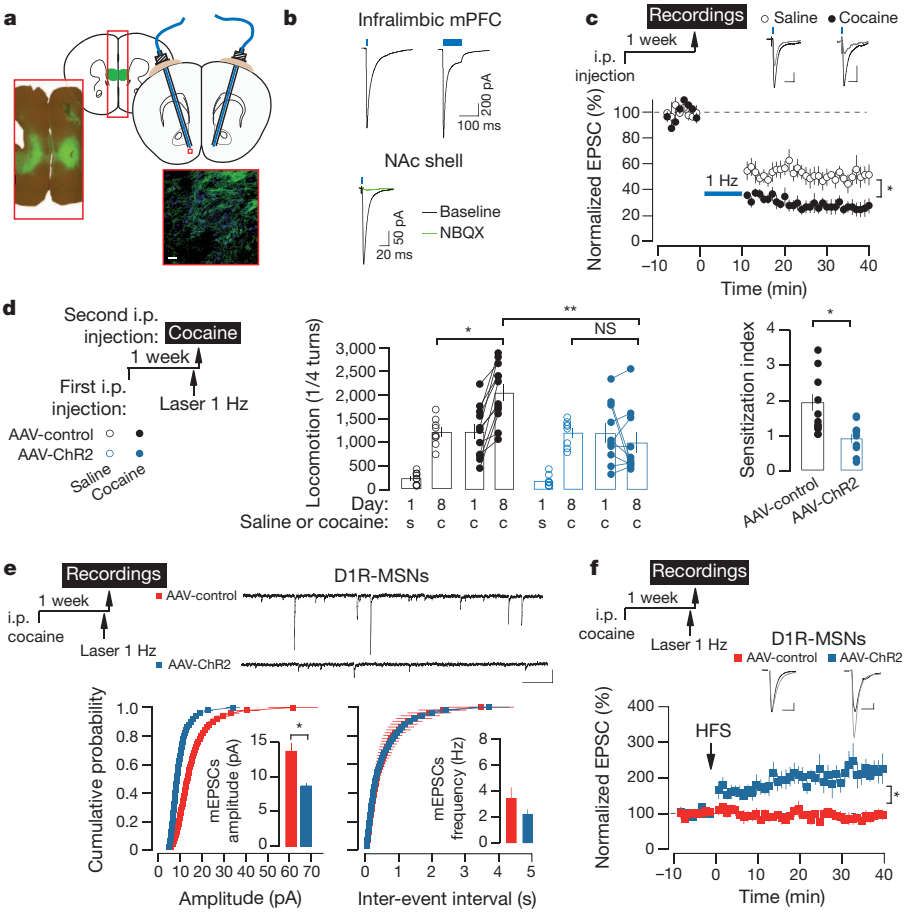


Figure 3 | Reversal of cocaine-evoked potentiation abolishes locomotor sensitization to cocaine. **a**, Schematic illustration of the site infected with AAV coding for eGFP-ChR2 and of the bilateral cannula implantation with optic fibres inserted in the nucleus accumbens shell. Insets show AAV-eGFP-ChR2 expression in the infralimbic cortex and high-resolution image of axonal inputs onto nucleus accumbens shell neurons (blue stains nuclei, ×60 magnification). **b**, Averaged current traces in response to light pulses (470 nm) of 4 ms and 100 ms in infected cells in slices of the infralimbic medial prefrontal cortex (mPFC; top). In nucleus accumbens (NAc) slices, light pulses (4 ms) evoked AMPAR EPSCs (blocked by NBQX 20 μM, green trace, bottom). **c**, Low-frequency stimulation (light pulses of 4 ms at 1 Hz for 10 min) induced depression of AMPAR EPSCs evoked with light in nucleus accumbens slices from mice injected with saline or cocaine one week previously (48 ± 5.7% and 27 ± 4.5%, respectively, $t_{14} = 2.851$). $n = 8$, $^*P < 0.05$. **d**, Quarter turns done by mice in the circular corridor for 60 min after injection represent locomotion (scatter plots of individual score and bars of mean ± s.e.m. are shown). Multiple-way repeated-measures analysis of variance for matching data yielded: interaction between day, virus and treatment $F_{(1,35)} = 13.41$, $P < 0.001$; effect of day $F_{(1,35)} = 78.16$, $P < 0.001$; effect of virus $F_{(1,35)} = 17.78$, $P < 0.001$; effect of treatment $F_{(1,35)} = 88.65$, $P < 0.001$. Wilcoxon or *t*-test: $^*P < 0.001$ for cocaine versus saline pre-treatment on day 8 in AAV-control and $P = 0.203$ cocaine versus saline pre-treatment on day 8 in AAV-ChR2 infected mice; $^{**}P = 0.0015$ for locomotor response to cocaine on day 8 for AAV-ChR2 versus AAV-control infected mice). No effect of light protocol on locomotor response to cocaine on day 8 when mice receive saline on day 1 was detected ($P > 0.05$ AAV-ChR2 versus AAV control). The sensitization index (see Methods) was significantly reduced in AAV-ChR2 infected mice treated with the light stimulation protocol (4 ms pulse, 1 Hz, 10 min) 45 min before the cocaine injection on day 8 ($t_{20} = 3.47$). $n = 8$–12, $^*P < 0.05$. NS, not significant. **e**, Acute slices of cocaine- and laser-treated AAV-ChR2 or AAV-control infected mice were prepared on day 8 (45 min after light stimulation). Sample traces, cumulative probability and mean values of amplitude and frequency for mEPSCs recorded from D1R-MSNs are shown. mEPSC amplitude was decreased in AAV-ChR2 compared to AAV-control infected mice ($t_{17} = 5.08$; Kolmogorov–Smirnov: $P < 0.001$). mEPSC frequency was unaltered ($t_{17} = 1{,}45$; Kolmogorov–Smirnov: $P > 0.05$). Scale bars: 200 ms and 20 pA. $n = 9$–10, $^*P < 0.05$. **f**, The light stimulation protocol restored HFS LTP in D1R-MSNs from AAV-ChR2 but not AAV-control infected mice (191.0 ± 21.2% and 93.0 ± 9.1%, respectively, $t_{12} = 4.26$). Scale bars: 10 ms and 20 pA. $n = 7$, $^*P < 0.05$. Error bars show s.e.m.

We next applied the above-validated protocol *in vivo* with the goal of establishing a causal link between cocaine-evoked plasticity and behavioural sensitization. We placed light guides into the ventral striatum, past the nucleus accumbens core, thus preferentially aiming at the principal cortical input onto MSNs of the nucleus accumbens shell, that is, axons that have their origin in the infralimbic cortex (Supplementary Fig. 8). When freely moving mice were treated with the optogenetic depotentiation protocol 45 min before the injection of the cocaine challenge at day 8, locomotor sensitization was completely erased. In control experiments, light stimulation did not affect the locomotor response at day 8 when the first injection was saline instead of cocaine or when a control virus was used (Fig. 3d). To ensure that the light stimulation restored normal transmission, we recorded *ex vivo* mEPSCs and found that an effective light treatment significantly reduced their mean amplitude in D1R-MSNs (Fig. 3e) to a level comparable to baseline transmission (Fig. 1e) without any effect on paired pulse ratio (Supplementary Fig. 10). In contrast, the mEPSC amplitude and frequency were not modified in D2R-MSNs after cocaine (Supplementary Fig. 11). Lastly, after cocaine treatment and optogenetic depotentiation the HFS protocol resulted in LTP *ex vivo* (Fig. 3f), confirming the restoration of baseline transmission in D1R-MSNs.

Behavioural sensitization typically refers to the observation of increased locomotor responses with repeated injections of cocaine. However, sensitization becomes apparent already after the second injection and is best observed about a week later[22] (Fig. 3). Although certainly not sufficient to induce addiction, such early forms of drug-induced adaptations are considered permissive building blocks for more definite behavioural alterations. If this is the case, the synaptic

©2011 Macmillan Publishers Limited. All rights reserved

potentiation onto D1R-MSNs should still be observed after chronic cocaine exposure, for example after 5 daily injections followed by a 10-day withdrawal. To confirm this hypothesis we recorded mEPSCs at the end of this protocol, as well as 1 h and 24 h after a challenge injection (Supplementary Fig. 12). This experiment confirmed that excitatory synapses onto D1R-MSNs were selectively potentiated at the end of the withdrawal period.

We then tested whether, after this chronic exposure to cocaine, light stimulation was still effective in reversing locomotor sensitization. After the 5 days of injection we imposed 10 days of withdrawal, and treated the mice with light stimulation 45 min before injecting a challenge dose. This completely reversed locomotor sensitization (Fig. 4a). When a second challenge injection was administered 24 h later, sensitized responses were again only observed in control mice. As a control, we did not observe an acute effect on locomotor behaviour during light stimulation (Supplementary Fig. 14).

Lastly, to estimate how long the effect of the light stimulation lasted, we tested for behavioural sensitization 5 days after the intervention (Fig. 4b). No significant locomotor sensitization was observed even when the mice were challenged at this extended time after light stimulation.

Our results identify NMDAR- and ERK-dependent LTP in D1R-MSNs of the nucleus accumbens as a form of synaptic plasticity required for locomotor sensitization to cocaine. ERK activation probably constitutes a general feature of addictive drugs, because in all brain regions that receive dopamine inputs tetrahydrocannabinol, amphetamines, morphine and nicotine also activate ERK signalling[4]. Through cocaine-driven ERK phosphorylation, potentiation is induced selectively in D1R-MSNs, which leads to the occlusion of HFS-driven LTP. These findings are in line with observations that ERK activation may control AMPAR trafficking directly, an effect that may also be maintained over days through activation of ERK nuclear targets leading to gene regulation[23].

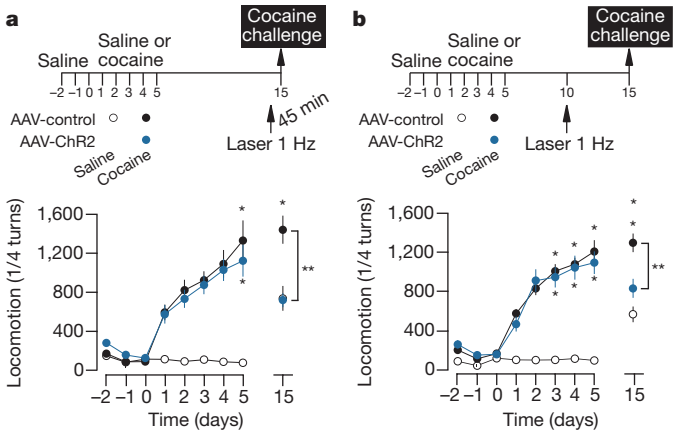


Figure 4 | Optogenetic depotentiation resets behavioural sensitization induced by chronic cocaine injections. **a**, After 5 daily cocaine injections a robust locomotor sensitization is measured in AAV-control and AAV-ChR2 infected mice. The light stimulation protocol (45 min before a challenge injection of cocaine on day 15) abolished the locomotor sensitized response in AAV-ChR2 infected mice (one-way analysis of variance: effect of treatment $F_{(2,41)} = 88.65$, $P < 0.001$. Post-hoc comparison by Bonferroni test yielded: $^{*}P < 0.001$ for cocaine versus saline pre-treatment in AAV control; $^{**}P < 0.001$ for AAV-ChR2 versus AAV-control infected mice, pre-treated with cocaine ($n = 13$–15). **b**, Light stimulation protocol applied on day 10 still reduced the locomotor sensitized response to a cocaine challenge injection on day 15 (one-way analysis of variance: effect of treatment $F_{(2,45)} = 15.31$, $P < 0.001$. Post-hoc comparison by Bonferroni test yielded: $^{*}P < 0.001$ for cocaine versus saline pre-treatment in AAV-control infected mice; $^{**}P = 0.003$ for AAV-ChR2 versus AAV-control infected mice, pre-treated with cocaine. No sensitization in AAV-ChR2 ($P = 0.27$, compared to saline pre-treatment; $n = 11$–21). Error bars show s.e.m.

We have explored and gained insight into the molecular mechanisms of synaptic adaptations to develop a strategy for reversal of cocaine-evoked potentiation with the goal of normalizing behaviour. We chose an optogenetic depotientiaton of inputs from the infralimbic cortex to the nucleus accumbens shell because of the strong anatomical connection and the functional implication of this projection in cocaine-seeking behaviour[24,25]. It is an appealing idea that behavioural adaptation, which closely reflects the potentiation of excitatory transmission onto D1R-MSNs, is due to an imbalance of the two classes of MSNs[26].

Several studies have already reported that pharmacological and molecular manipulations of key players in synaptic plasticity in the nucleus accumbens can affect adaptive behaviours associated with addictive drug exposure. For example, the inhibition of calcium-permeable AMPARs, a hallmark of late-stage cocaine-evoked synaptic plasticity in the nucleus accumbens[27] and the viral expression of a peptide that impairs GluA1 trafficking[28], reduce cue-induced cocaine seeking and cocaine-primed reinstatement, respectively.

We provide proof of principle that optogenetic manipulations can be used to reverse cocaine-evoked synaptic plasticity and thus abolish locomotor sensitization. Although light stimulation fully resets locomotor behaviour, sensitization begins to reappear after a few days, suggesting that several treatment sessions may be required to obtain long-lasting effects. This is not surprising, as chronic cocaine exposure also induces a number of additional adaptive changes including structural remodelling (for example, increase in spines[29]) and alterations of gene expression[23].

Sensitization to cocaine-associated stimuli has been linked with incentive saliency[30] and may explain the exceptionally strong motivation of addicts to obtain the drug. With chronic use, early adaptive changes such as those described here may build up to enhance craving during cocaine withdrawal[30]. Successful interventions that reverse these changes in animal models could inspire novel treatments for human addiction, a disease with a high social burden. Indeed, novel protocols of deep brain stimulation or transcranial magnetic stimulation may induce forms of synaptic plasticity that reverse drug-evoked adaptations, thus curbing the risk of relapse.

METHODS SUMMARY

All experiments were reviewed by the institutional ethics committee and approved by the relevant authorities of the Canton of Geneva. C57BL/6 or heterozygous BAC transgenic female and male mice, in which eGFP expression was driven by either D1R (*drd1a*-eGFP) or D2R (*drd2*-eGFP) gene regulatory elements were injected intraperitoneally with saline or cocaine. We then prepared brain slices for electrophysiological recordings as previously described[3] or placed the mice in an apparatus to quantify locomotor behaviour. Locomotor sensitization or synaptic plasticity was monitored at the various time points after the cocaine injection. A two-injection or five-injection protocol to induce locomotor sensitization was used as previously described[22]. Standard surgical procedures[3] were used to infect mice with ChR2-AAV or a control AAV (0.5 µl) in the infralimbic medial prefrontal cortex while the light guides were aimed at both nucleus accumbens (shell). In depotentiation experiments, a 473-nm solid-state laser was used to carry out the *in vivo* light stimulation protocol in awake mice (600 pulses of 4 ms duration at 1 Hz, 10–20 mW), 45 min or 5 days before behavioural testing or *ex vivo* electrophysiology recordings, respectively.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 June; accepted 10 November 2011.
Published online 7 December 2011.

1. Lüscher, C. & Malenka, R. C. Drug-evoked synaptic plasticity in addiction: from molecular changes to circuit remodeling. *Neuron* **69,** 650–663 (2011).
2. Chen, R. *et al.* Abolished cocaine reward in mice with a cocaine-insensitive dopamine transporter. *Proc. Natl Acad. Sci. USA* **103,** 9333–9338 (2006).
3. Brown, M. T. *et al.* Drug-driven AMPA receptor redistribution mimicked by selective dopamine neuron stimulation. *PLoS ONE* **5,** e15870 (2010).
4. Valjent, E., Pages, C., Herve, D., Girault, J. A. & Caboche, J. Addictive and non-addictive drugs induce distinct and specific patterns of ERK activation in mouse brain. *Eur. J. Neurosci.* **19,** 1826–1836 (2004).

©2011 Macmillan Publishers Limited. All rights reserved

5. Pascoli, V. *et al.* Cyclic adenosine monophosphate-independent tyrosine phosphorylation of NR2B mediates cocaine-induced extracellular signal-regulated kinase activation. *Biol. Psychiatry* **69,** 218–227 (2011).
6. Lu, L., Koya, E., Zhai, H., Hope, B. T. & Shaham, Y. Role of ERK in cocaine addiction. *Trends Neurosci.* **29,** 695–703 (2006).
7. Girault, J. A., Valjent, E., Caboche, J. & Herve, D. ERK2: a logical AND gate critical for drug-induced plasticity? *Curr. Opin. Pharmacol.* **7,** 77–85 (2007).
8. Thomas, M. J., Kalivas, P. W. & Shaham, Y. Neuroplasticity in the mesolimbic dopamine system and cocaine addiction. *Br. J. Pharmacol.* **154,** 327–342 (2008).
9. Bertran-Gonzalez, J. *et al.* Opposing patterns of signaling activation in dopamine D_1 and D_2 receptor-expressing striatal neurons in response to cocaine and haloperidol. *J. Neurosci.* **28,** 5671–5685 (2008).
10. Kramer, P. F. *et al.* Dopamine D2 receptor overexpression alters behavior and physiology in *Drd2-EGFP* mice. *J. Neurosci.* **31,** 126–132 (2011).
11. Lammel, S., Ion, D. I., Roeper, J. & Malenka, R. C. Projection-specific modulation of dopamine neuron synapses by aversive and rewarding stimuli. *Neuron* **70,** 855–862 (2011).
12. Wolf, M. E. The Bermuda Triangle of cocaine-induced neuroadaptations. *Trends Neurosci.* **33,** 391–398 (2010).
13. Kourrich, S., Rothwell, P. E., Klug, J. R. & Thomas, M. J. Cocaine experience controls bidirectional synaptic plasticity in the nucleus accumbens. *J. Neurosci.* **27,** 7921–7928 (2007).
14. Boudreau, A. C. & Wolf, M. E. Behavioral sensitization to cocaine is associated with increased AMPA receptor surface expression in the nucleus accumbens. *J. Neurosci.* **25,** 9144–9151 (2005).
15. Boudreau, A. C., Reimers, J. M., Milovanovic, M. & Wolf, M. E. Cell surface AMPA receptors in the rat nucleus accumbens increase during cocaine withdrawal but internalize after cocaine challenge in association with altered activation of mitogen-activated protein kinases. *J. Neurosci.* **27,** 10621–10635 (2007).
16. English, J. D. & Sweatt, J. D. A requirement for the mitogen-activated protein kinase cascade in hippocampal long term potentiation. *J. Biol. Chem.* **272,** 19103–19106 (1997).
17. Xie, G. Q. *et al.* Ethanol attenuates the HFS-induced, ERK-mediated LTP in a dose-dependent manner in rat striatum. *Alcohol. Clin. Exp. Res.* **33,** 121–128 (2009).
18. Shen, W., Flajolet, M., Greengard, P. & Surmeier, D. J. Dichotomous dopaminergic control of striatal synaptic plasticity. *Science* **321,** 848–851 (2008).
19. Zhu, J. J., Qin, Y., Zhao, M., Van Aelst, L. & Malinow, R. Ras and Rap control AMPA receptor trafficking during synaptic plasticity. *Cell* **110,** 443–455 (2002).
20. Valjent, E. *et al.* Regulation of a protein phosphatase cascade allows convergent dopamine and glutamate signals to activate ERK in the striatum. *Proc. Natl Acad. Sci. USA* **102,** 491–496 (2005).
21. Pierce, R. C., Pierce-Bancroft, A. F. & Prasad, B. M. Neurotrophin-3 contributes to the initiation of behavioral sensitization to cocaine by activating the Ras/mitogen-activated protein kinase signal transduction cascade. *J. Neurosci.* **19,** 8685–8695 (1999).
22. Valjent, E. *et al.* Mechanisms of locomotor sensitization to drugs of abuse in a two-injection protocol. *Neuropsychopharmacology* **35,** 401–415 (2010).
23. Brami-Cherrier, K., Roze, E., Girault, J. A., Betuing, S. & Caboche, J. Role of the ERK/MSK1 signalling pathway in chromatin remodelling and brain responses to drugs of abuse. *J. Neurochem.* **108,** 1323–1335 (2009).
24. Sesack, S. R., Deutch, A. Y., Roth, R. H. & Bunney, B. S. Topographical organization of the efferent projections of the medial prefrontal cortex in the rat: an anterograde tract-tracing study with *Phaseolus vulgaris* leucoagglutinin. *J. Comp. Neurol.* **290,** 213–242 (1989).
25. LaLumiere, R. T., Niehoff, K. E. & Kalivas, P. W. The infralimbic cortex regulates the consolidation of extinction after cocaine self-administration. *Learn. Mem.* **17,** 168–175 (2010).
26. Lobo, M. K. *et al.* Cell type-specific loss of BDNF signaling mimics optogenetic control of cocaine reward. *Science* **330,** 385–390 (2010).
27. Conrad, K. L. *et al.* Formation of accumbens GluR2-lacking AMPA receptors mediates incubation of cocaine craving. *Nature* **454,** 118–121 (2008).
28. Anderson, S. M. *et al.* CaMKII: a biochemical bridge linking accumbens dopamine and glutamate systems in cocaine seeking. *Nature Neurosci.* **11,** 344–353 (2008).
29. Robinson, T. E. & Kolb, B. Structural plasticity associated with exposure to drugs of abuse. *Neuropharmacology* **47** (suppl. 1), 33–46 (2004).
30. Vanderschuren, L. J. & Pierce, R. C. Sensitization processes in drug addiction. *Curr. Top. Behav. Neurosci.* **3,** 179–195 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Kalivas and the members of the Lüscher laboratory for suggestions on the manuscript. This work is supported by the Swiss National Science Foundation (C.L.) and ''Synapsy'', a National Competence Center in Research (NCCR) of the Swiss Confederation on the synaptic basis of mental disorders.

Author Contributions V.P. carried out all electrophysiology experiments and was helped by M.T. with the behavioural experiments. C.L. designed the study and wrote the manuscript together with V.P. and M.T.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.L. (christian.luscher@unige.ch).

©2011 **Macmillan Publishers Limited. All rights reserved**

METHODS

Mice. C57BL/6 or heterozygous BAC transgenic mice, in which eGFP expression was driven by either D1R (*drd1a*-eGFP) or D2R (*drd2*-eGFP) gene regulatory elements were backcrossed[31] in C57Bl/6 mice for three to four generations, were used. Mice were housed in groups of 3–4 except for those implanted with guide cannulae, in which case animals were housed separately. All animals were kept in a temperature- and hygrometry-controlled environment with a 12 h light/12 h dark cycle. Mice were injected intraperitoneally with 20 mg kg^{-1} cocaine, 40 mg kg^{-1} SL327 (dissolved in 25% DMSO) or 0.9% saline (injection volume 10 ml kg^{-1}). Immediately after injection, mice were placed in the locomotor recording apparatus for 1 h. All procedures were approved by the Institutional Animal Care and Use Committee of the University of Geneva.

Locomotor sensitization. Locomotor activity was measured as the number of quarter turns entirely crossed by a mouse in a circular corridor. Locomotor chamber apparatus was placed under a video tracking system (Any-maze, Stoelting) and measurements were made automatically by the software. After 3 days of habituation to the test apparatus, mice underwent the experimental procedure, which consisted of two sessions of 60 min separated by 1 week (or 1 month), called day 1 and day 8 (or day 30). During the day 1 session, mice received saline or cocaine and were placed immediately in the corridor for 60 min. One week or 1 month later (day 8 or day 30), a second session was performed during which all mice were injected with cocaine before being placed in the circular corridor for 60 min. The light stimulation protocol (600 pulses of 4 ms at 1 Hz) was done 45 min before the second cocaine injection. To compare the effects of various times after the first injections or various virus infections, locomotor activity in response to the second cocaine injection was normalized to the mean locomotor activity of saline-pre-treated mice and the sensitization index was calculated by dividing the normalized locomotor response to the second injection by the normalized response to the first injection. Locomotor sensitization was also evaluated during challenge sessions that followed a chronic treatment (5 days of cocaine 15 mg kg^{-1}, 10 days withdrawal). The light stimulation protocol was done 45 min or 5 days before the challenge injection of cocaine.

Virus stereotaxic injection of ChR2-AAV or control AAV. AAV1 viruses produced at the University of North Carolina (Vector Core Facility) were injected into the infralimbic cortex of 15–20 g wild-type or BAC transgenic mice. Anaesthesia was induced and maintained with isoflurane (Baxter AG). The animal was placed in a stereotaxic frame (Angle One) and craniotomies were performed using stereotaxic coordinates (anterio-posterior, +1.9; medio-lateral, ±0.3; dorso-ventral, 2.4–2.6). Injections of AAV1 viruses (0.5 µl) were carried out using graduated pipettes (Drummond Scientific Company), broken back to a tip diameter of 10–15 µm, at a rate of ~0.05 µl min^{-1}. In all experiments the viruses were allowed a minimum of 3 weeks to incubate before any other procedures were carried out. As a control, some mice were injected with an AAV containing only GFP.

Cannula implantation. Following anaesthesia and craniotomy over the nucleus accumbens, two holes were drilled around the craniotomy and screws were placed in the holes. Two weeks after viral injections, guide cannulae (Plastics One) were lowered slowly into position using stereotaxic coordinates (bilaterally anterio-posterior, +1.5; medio-lateral, ±1.6; dorso-ventral 4.1; 15° angle) and cemented in place using dental cement (Lang Dental MFG Company) to encase the base of the guide cannulae and the screws. Once the cement had dried, a dummy cannula (Plastics One) was placed inside each guide cannula to prevent infection.

Slice electrophysiology. Coronal 200–250-µm slices of mouse forebrain were prepared in cooled artificial cerebrospinal fluid (ACSF) containing (in mM): NaCl 119, KCl 2.5, MgCl 1.3, $CaCl_2$ 2.5, Na_2HPO_4 1.0, $NaHCO_3$ 26.2 and glucose 11, bubbled with 95% O_2 and 5% CO_2. Slices were kept at 32–34 °C in a recording chamber superfused with 2.5 ml min^{-1} ACSF. Visualized whole-cell voltage-clamp recording techniques were used to measure holding and synaptic responses of MSNs of the nucleus accumbens shell, identified in some experiments by the presence of the GFP of BAC transgenic mice by using a fluorescent microscope (Olympus BX50WI, fluorescent light U-RFL-T). The holding potential was −70 mV, and the access resistance was monitored by a hyperpolarizing step of −14 mV with each sweep, every 10 s. Experiments were discarded if the access resistance varied by more than 20%. Synaptic currents were evoked by stimuli (50–100 µs) at 0.1 Hz through bipolar stainless steel electrodes placed at the cortex–nucleus-accumbens border. The internal solution contained (in mM): 140 k-gluconate, 5 KCl, 130 CsCl, 10 HEPES, 0.2 EGTA, 2 $MgCl_2$, 4 Na_2ATP, 0.3 Na_3GTP and 10 sodium creatine-phosphate. Currents were amplified (Multiclamp 700B, Axon Instruments), filtered at 5 kHz and digitized at 20 kHz (National Instruments Board PCI-MIO-16E4, Igor, WaveMetrics). The liquid junction potential was small (−3 mV), and therefore traces were not corrected. All experiments were carried out in the presence of picrotoxin (100 µM). LTP was induced by using the following HFS protocol: 100 pulses at 100 Hz repeated 4 times at 0.1 Hz paired with depolarization at 0 mV[32,33].

Miniature EPSCs were recorded in the presence of tetrodotoxin (0.5 µM). The frequency, amplitudes and kinetic properties of these currents were then analysed using the Mini Analysis software package (v.4.3, Synaptosoft). Cocaine or light illumination of ChR2-induced changes in cumulative miniature EPSC amplitude and inter-event interval distribution were analysed for statistical significance using the nonparametric two-sample Kolmogorov–Smirnov test (KyPlot) with a conservative critical probability level of $P < 0.05$.

Paired pulse ratio (PPR) was calculated by dividing the second evoked EPSC by the first with a 50-ms interval in between.

AMPAR EPSCs evoked with ChR2 stimulation by 4-ms light pulses (LED, Thorlabs) were recorded in the same conditions as electrically evoked synaptic currents. Low-frequency stimulation (LFS; 1 Hz for 10 min) was applied with light pulses and the magnitude of LTD was determined by comparing average EPSCs that were recorded 20–30 min after induction to EPSCs recorded immediately before induction.

***In vivo* stimulation of infralimbic cortex projections in the nucleus accumbens shell.** Virus injected and cannulated animals were allowed a minimum of 1 week to recover and to express the virus. 473-nm solid-state lasers (GMP, CH) were used to carry out the *in vivo* stimulation protocol in awake mice. A fibre optic (Thorlabs) was customized to enable the mouse to move freely during stimulation. Briefly, the plastic cap of a dummy cannula (Plastics One) was hollowed out and a hole of sufficient diameter for the fibre optic to pass through made in the top. This was threaded onto the fibre optic, one end of which was stripped to leave a 200 µm external diameter. The fibre was then lowered into the guide cannula on the mouse and the hollowed-out dummy cannula cap screwed onto the guide cannula. An FC/PC rotative fibre-optic rotary joint (Doric lenses) was used to release torsion in the fibre caused by the animal's rotation. The fibre was connected to the laser, which delivered 4 ms pulses at 1 Hz for 10 min, an established LTD protocol at excitatory synapses in the nucleus accumbens[34]. All stimulations were carried out in the mouse home cage (except in the experiment shown in Supplementary Fig. 10, in which stimulation was performed during locomotor recordings in the circular corridor) 45 min or 5 days before behavioural testing or *ex vivo* electrophysiology recordings.

31. Mameli, M. *et al.* Cocaine-evoked synaptic plasticity: persistence in the VTA triggers adaptations in the nucleus accumbens. *Nature Neurosci.* **12,** 1036–1041 (2009).
32. Gong, S. *et al.* Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. *J. Neurosci.* **27,** 9817–9823 (2007).
33. Kombian, S. B. & Malenka, R. C. Simultaneous LTP of non-NMDA- and LTD of NMDA-receptor-mediated responses in the nucleus accumbens. *Nature* **368,** 242–246 (1994).
34. Pennartz, C. M., Ameerun, R. F., Groenewegen, H. J. & Lopes da Silva, F. H. Synaptic plasticity in an *in vitro* slice preparation of the rat nucleus accumbens. *Eur. J. Neurosci.* **5,** 107–117 (1993).

©2011 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature10715

Modulation of TRPA1 thermal sensitivity enables sensory discrimination in *Drosophila*

Kyeongjin Kang[1]*, Vincent C. Panzano[1]*, Elaine C. Chang[1], Lina Ni[1], Alexandra M. Dainis[1], Adam M. Jenkins[2], Kimberly Regna[2], Marc A. T. Muskavitch[2,3,4] & Paul A. Garrity[1]

Discriminating among sensory stimuli is critical for animal survival. This discrimination is particularly essential when evaluating whether a stimulus is noxious or innocuous. From insects to humans, transient receptor potential (TRP) channels are key transducers of thermal, chemical and other sensory cues[1,2]. Many TRPs are multimodal receptors that respond to diverse stimuli[1–3], but how animals distinguish sensory inputs activating the same TRP is largely unknown. Here we determine how stimuli activating *Drosophila* TRPA1 are discriminated. Although *Drosophila* TRPA1 responds to both noxious chemicals[4] and innocuous warming[5], we find that TRPA1-expressing chemosensory neurons respond to chemicals but not warmth, a specificity conferred by a chemosensory-specific TRPA1 isoform with reduced thermosensitivity compared to the previously described isoform. At the molecular level, this reduction results from a unique region that robustly reduces the channel's thermosensitivity. Cell-type segregation of TRPA1 activity is critical: when the thermosensory isoform is expressed in chemosensors, flies respond to innocuous warming with regurgitation, a nocifensive response. TRPA1 isoform diversity is conserved in malaria mosquitoes, indicating that similar mechanisms may allow discrimination of host-derived warmth—an attractant—from chemical repellents. These findings indicate that reducing thermosensitivity can be critical for TRP channel functional diversification, facilitating their use in contexts in which thermal sensitivity can be maladaptive.

Highly temperature-responsive TRP cation channels, thermoTRPs, mediate thermosensation from insects to mammals[1,2] and are important for human pain and inflammation[6]. Like mammalian thermoTRPs, *Drosophila melanogaster* TRPA1 is both a thermal and chemical sensor, responding to innocuous warmth (above ~25–27 °C)[5,7] and noxious chemicals[4]. TRPA1 acts in thermosensors within the brain to modulate thermal preference over 18–32 °C (ref. 5), innocuous temperatures compatible with fly survival[8], and in gustatory chemosensors to inhibit ingestion of electrophiles[4], reactive chemicals like allyl isothiocyanate (AITC, found in wasabi) and *N*-methylmaleimide (NMM) that rapidly incapacitate flies (Supplementary Fig. 1). The responsiveness of TRPA1 to both innocuous and noxious stimuli raises the question of how these stimuli are distinguished to elicit distinct behavioural responses. Mammals face similar issues; for example, TRPM8 transduces both innocuous and noxious cold[1–3].

We previously reported TRPA1-expressing chemosensors in the labral sense organ[4]; using improved immunostaining conditions, we now also detect specific TRPA1 protein expression in labellar chemosensors (Fig. 1a, b). Extracellular tip recording[9] indicated these neurons were TRPA1-dependent chemosensors; they responded to the electrophile NMM with robust spiking in wild type but not *TrpA1* mutants

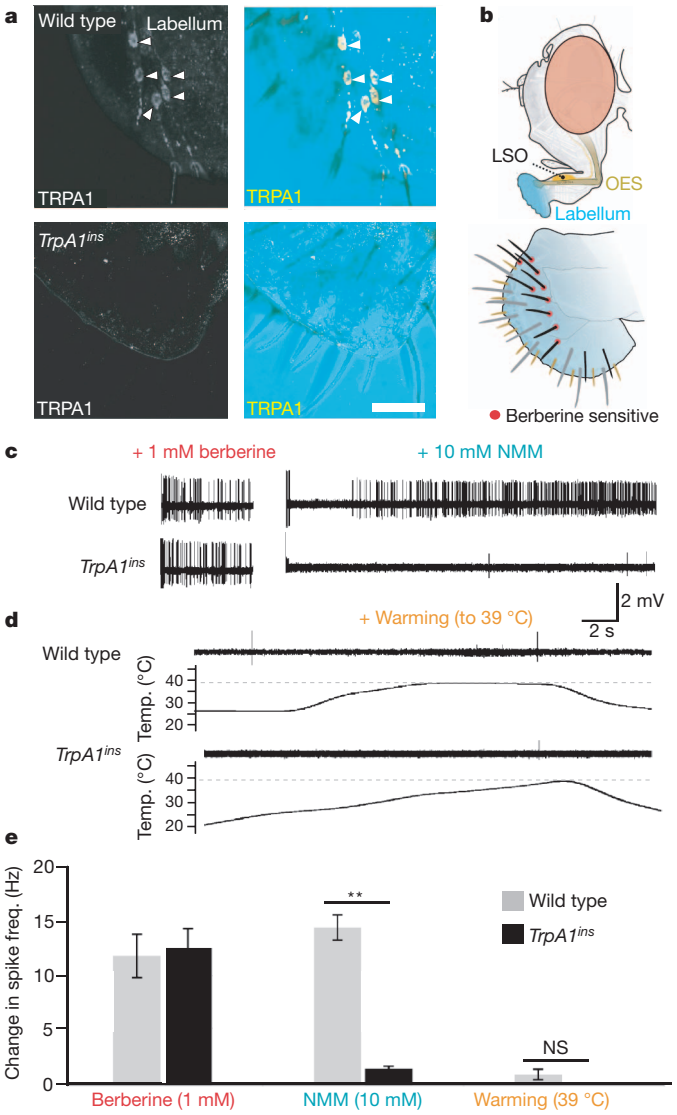


Figure 1 | TRPA1-dependent gustatory neurons do not respond to heat. **a**, TRPA1 immunostaining of wild-type (top) and *TrpA1*^{ins} (bottom) labella. Right, differential interference contrast overlay of labellar structures. Arrowheads, chemosensor cell bodies. Scale bar, 20 μm. **b**, Top, *Drosophila* gustatory organs: LSO, labral sense organ; OES, oesophagus. Bottom, labellar bristles. Brown, s-type; grey, L-type; black, i-type; berberine-sensitive bristles were targeted for electrophysiology. **c**, **d**, Bristle responses to: berberine (1 mM) and NMM (10 mM) (**c**); warming (**d**). Temp., temperature. **e**, Average spike rate after subtracting electrolyte-only baseline. freq., frequency. $^{**}P < 0.01$; NS, not significant ($P > 0.05$), *t*-test. All data are mean ± standard error of the mean (s.e.m.). Warming reached an average maximum temperature of 39.0 ± 0.6 °C (mean ± standard deviation).

[1]National Center for Behavioral Genomics and Volen Center for Complex Systems, Department of Biology, Brandeis University, Waltham, Massachusetts 02454, USA. [2]Biology Department, Boston College, Newton, Massachusetts 02467, USA. [3]Broad Institute, Cambridge, Massachusetts 02142, USA. [4]Harvard School of Public Health, Boston, Massachusetts 02115, USA.
*These authors contributed equally to this work.

(Fig. 1c and Supplementary Fig. 2). The mutant defect was electrophile specific, as *TrpA1* mutants responded like wild type to berberine chloride (Fig. 1c), a bitter compound that also activates these neurons[10]. In contrast, warming to ~39 °C, from innocuously warm to the noxious range, elicited no spiking in these cells (Fig. 1d, e). This is notable as the effectiveness of TRPA1 in conferring warmth sensitivity has led to its use as a thermogenetic tool[5,11]. Thus, despite the known sensitivity of TRPA1 to both temperature and chemicals, these chemosensors are warmth insensitive.

In addition to the previously characterized transcript, *TrpA1(B)*, a transcript with an alternative 5′ end, *TrpA1(A)*, has been annotated[12] (Fig. 2a). These transcripts encode protein isoforms with distinct amino termini, but the same ankyrin and transmembrane domains (Fig. 2b). Polymerase chain reaction with reverse transcription (RT–PCR) demonstrated differential expression: *TrpA1(A)* was expressed in the proboscis, which houses the TRPA1-expressing chemosensors, whereas *TrpA1(B)* predominated elsewhere in the head, where TRPA1-expressing thermosensors are located (Fig. 2c).

Examined in *Xenopus* oocytes, TRPA1(A) was much less thermosensitive than TRPA1(B), as reflected in its temperature coefficient (Q_{10}), the fold change in current per 10 °C change[1,2]. Arrhenius plot analysis[13] yielded a Q_{10} of ~9 for TRPA1(A) versus ~116 for TRPA1(B) (Fig. 2d–f). In addition, whereas TRPA1(B) was essentially inactive at low temperatures, TRPA1(A)-dependent currents were observed ≤15 °C, further reducing the temperature-dependent activity differential of TRPA1(A) (Supplementary Fig. 3). The maximum heat-activated current for TRPA1(A) was also significantly lower (Supplementary Fig. 4). Lastly, the transition (or threshold) temperature for increased temperature responsiveness was 29.7 ± 0.3 °C for TRPA1(A) versus 27.8 ± 0.4 °C for TRPA1(B) ($P < 0.01$, *t*-test). As the innocuous warm temperature range in *Drosophila* is of particular behavioural relevance, the Q_{10} from 27–37 °C (below the ~38 °C nociceptive threshold in *Drosophila*[14]) was also calculated, yielding 6.2 ± 0.5 for TRPA1(A) and 90 ± 8 for TRPA1(B) (Fig. 2f). Other properties were largely unaffected; both channels responded robustly to electrophiles and had similar voltage sensitivities (Fig. 2g, h). TRPA1(A) and TRPA1(B) had similar maximum current amplitudes at 300 μM NMM, with half-maximum effective concentration (EC_{50}) values of 176 ± 12 and 128 ± 9 μM, respectively (Fig. 2i).

The reduced thermosensitivity of TRPA1(A) could account for the chemosensors' warmth insensitivity. But although TRPA1(A) is less temperature sensitive than TRPA1(B), its Q_{10} resembles several TRPs implicated in warmth sensitivity[15–17]. To assess whether TRPA1(A) could confer warmth sensitivity upon *Drosophila* chemosensors, each isoform was used to rescue a *TrpA1* mutant. We previously demonstrated that expressing TRPA1(B) in TRPA1-dependent chemosensors using Gr66a-Gal4 rescues the *TrpA1* mutant behavioural defect[4]. Using electrophysiology, we found both isoforms restored NMM responsiveness (Fig. 3a and Supplementary Fig. 5a, b), but only TRPA1(B) conferred warmth sensitivity (Fig. 3b, c). These differences did not require properties unique to TRPA1-dependent chemosensors. Each isoform was expressed ectopically in sweet-responsive chemosensors using Gr5a-Gal4 (ref. 18). Both isoforms conferred electrophile sensitivity upon these normally electrophile-insensitive neurons, but only TRPA1(B) conferred thermosensitivity (Fig. 3d–f and Supplementary Fig. 4c, d). The inability of TRPA1(A) to confer warmth sensitivity on fly chemosensors emphasizes that although a Q_{10} above 5 makes TRPA1(A) more thermally sensitive than most ion channels, *in vivo* testing is important in evaluating whether a channel is sufficiently thermosensitive to make a specific neuron warmth responsive.

These data support a model in which the specificity of TRPA1-expressing gustatory neurons for chemicals is established by their selective expression of TRPA1(A), an isoform unable to confer warmth sensitivity. In contrast, the chemical sensitivity of TRPA1(B) should render TRPA1-dependent thermosensors sensitive to reactive chemicals. However, the location of TRPA1-dependent anterior cell thermosensors

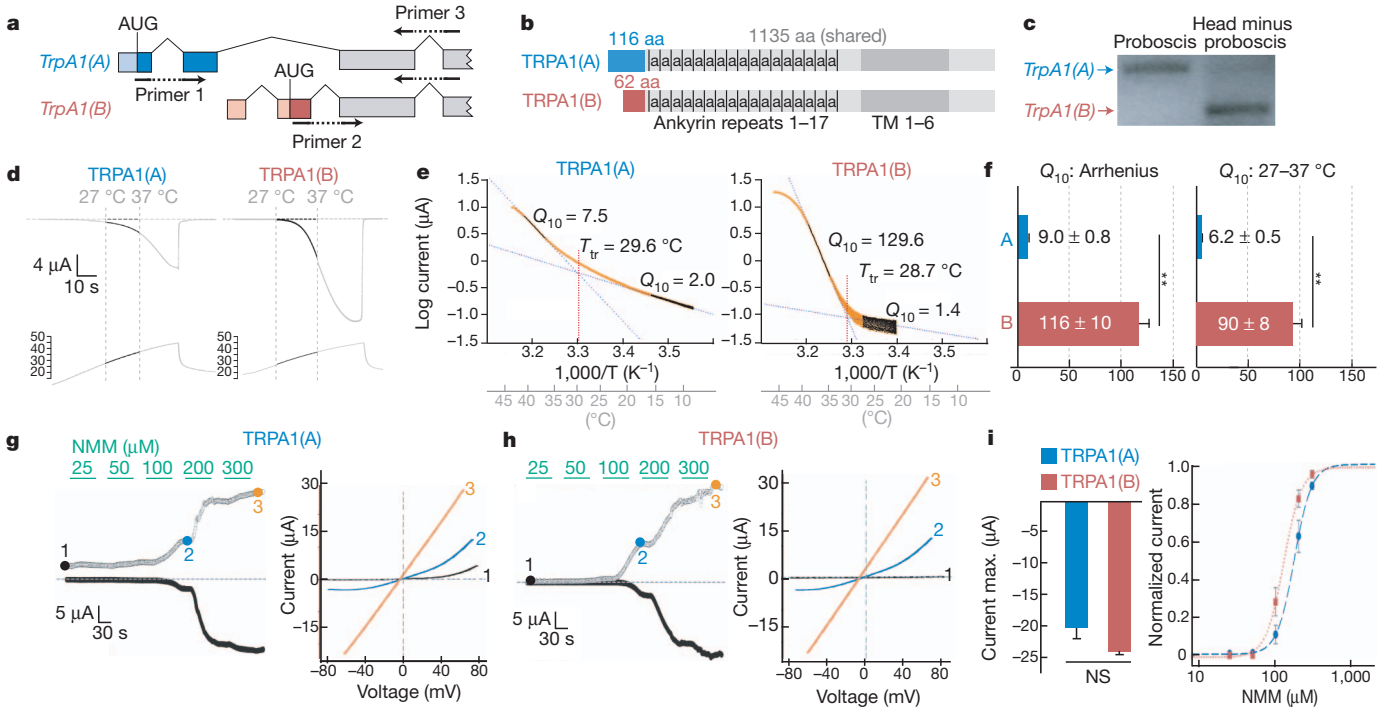


Figure 2 | TRPA1 isoform diversity yields tissue-specific channels with different thermal sensitivities. **a**, *TrpA1* gene structure and primer locations. **b**, TRPA1 proteins. Dark red and blue boxes denote isoform-specific sequences. Dark grey, transmembrane (TM) region. a, ankyrin repeat; aa, amino acids. **c**, RT–PCR analysis of dissected tissue. **d**, **e**, TRPA1(A)- and TRPA1(B)-dependent currents (**d**) and Arrhenius plots (**e**) in oocytes. T_{tr}, transition temperature. **f**, Q_{10} values from Arrhenius plot (left) or 27–37 °C (right). **g**, **h**, Left panels, NMM responsiveness of TRPA1(A) (**g**) and TRPA1(B) (**h**). Right panels, *I–V* relationships at points marked at left. **i**, Mean amplitudes at 300 μM NMM (left) and NMM dose–response (right). All data, mean ± s.e.m. $^{**}P < 0.01$; NS, not significant, *t*-test.

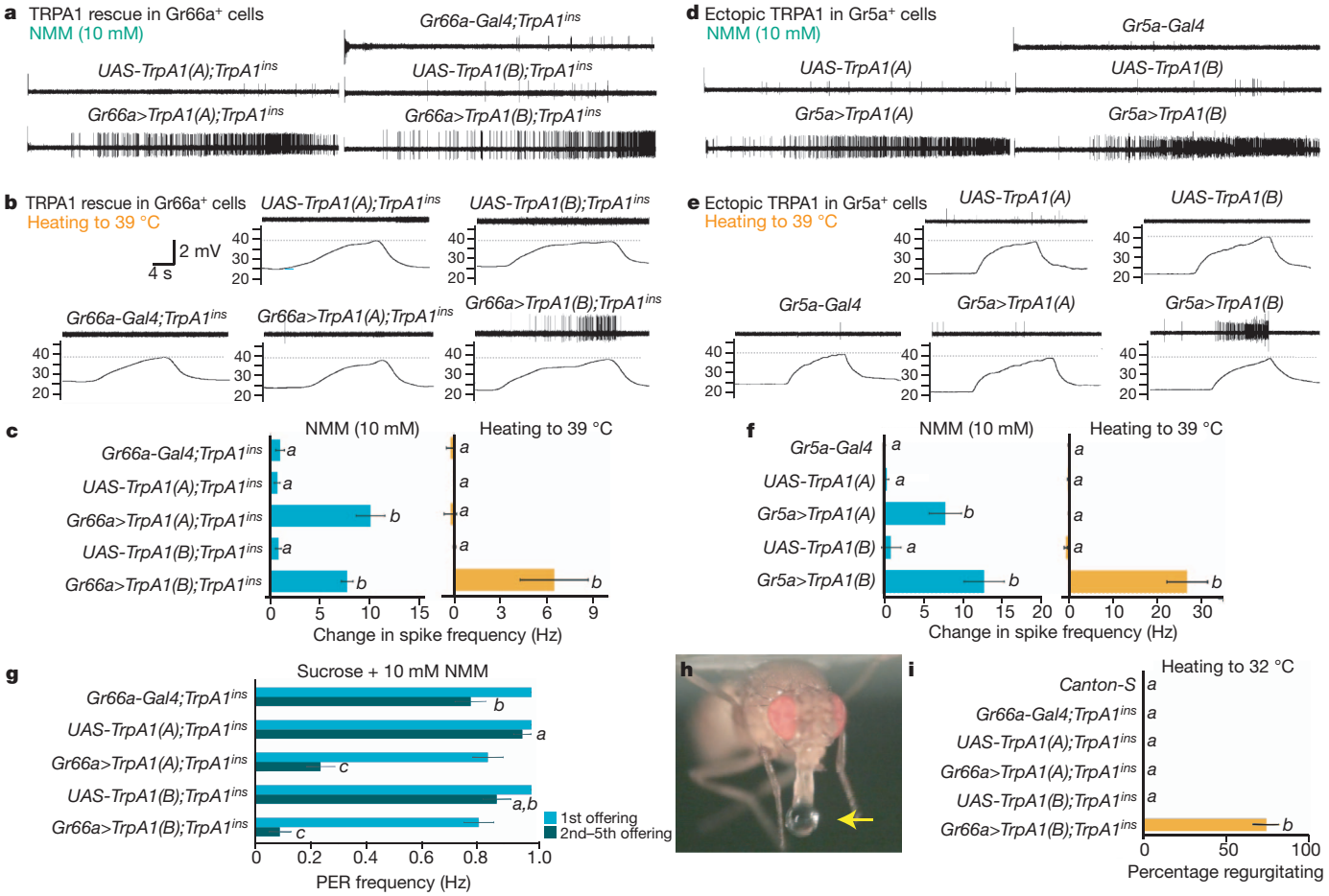


Figure 3 | TRPA1 isoform diversity determines sensory specificity of gustatory neurons. **a**–**c**, *TrpA1* mutant, berberine-sensitive i-type bristles expressing different TRPA1 isoforms. **a**, **b**, Electrophysiological responses to NMM (**a**) and warming (**b**). **c**, Quantification. **d**–**f**, L-type bristles expressing TRPA1 isoforms. Responses to NMM (**d**) and warming (**e**). **f**, Quantification. **g**, Rescue of *TrpA1* mutant behavioural response to NMM-containing food. PER, proboscis extension response. **h**, Warmth-induced regurgitation in *TrpA1* mutant rescued with TRPA1(B). **i**, Regurgitation upon warming from room temperature (~23 °C) to 32 °C. In **c**, **f**, **g** and **i**, statistically distinct groups are marked by *a*, *b* and *c*. (Tukey honestly significant difference (HSD) test, $\alpha = 0.01$). Data are mean ± s.e.m.

inside the head[5] should minimize exposure to environmental irritants. Interestingly, multiple TRPV1 and TRPM1 isoforms are present in humans and other mammals[2,17,19,20], suggesting the potential generality of isoform diversity in modulating TRP functions.

The behavioural significance of discriminating noxious from innocuous TRPA1 activators was examined by testing gustatory responses of *TrpA1* mutants rescued by chemosensor expression of each isoform. *TrpA1* mutants exhibit decreased avoidance of reactive-electrophile-containing food[4]. Each isoform rescued this behaviour (Fig. 3g). However, TRPA1(B) also triggered a nocifensive response to innocuous warming. When allowed to ingest water to satiation and warmed to ~32 °C, neither wild-type nor TRPA1(A) rescue animals showed detectable gustatory responses (Fig. 3h, i). However, warming TRPA1(B) rescue flies caused ~75% to regurgitate (Fig. 3h, i and Supplementary Movie 1). Thus, substituting TRPA1(B) for TRPA1(A) in chemosensors disrupts discrimination of noxious from innocuous stimuli and demonstrates the negative behavioural consequence of misregulated thermosensitivity.

To probe how the alternative N termini in TRPA1 confer distinct properties, conserved residues within these regions were mutated (Fig. 4a). Mutating either a cysteine (C105) or two basic residues (R113, R116) in TRPA1(A) markedly increased temperature responsiveness (Fig. 4a–c and Supplementary Figs 6 and 7). Whereas the Q_{10} of wild-type TRPA1(A) was <10, the TRPA1(A) mutants exhibited Q_{10} values of >50 (Fig. 4b, c), greater than the reported Q_{10} values of canonical thermoTRPs like TRPM8 (~24)[21] and TRPV1 (~40)[22]. In addition, the TRPA1(A) mutants conducted little current below the threshold, increasing the temperature-dependent activity differential (Fig. 4c and Supplementary Fig. 8). The enhanced sensitivities of the mutants seemed to be temperature specific, as NMM sensitivity was not increased (Supplementary Fig. 8). These data indicate that TRPA1(A) retains all the requirements for robust thermosensation, but contains a modulatory region preventing those elements from exerting their full effect.

For TRPA1(B), mutating either a conserved tryptophan or two basic residues in the N terminus yielded channels retaining robust thermosensitivity ($Q_{10} > 50$; Fig. 4b and Supplementary Fig. 6). The thresholds of the TRPA1(A) and TRPA1(B) mutants were all ~30–34 °C, within the innocuous warm range but above the ~28 °C value for wild-type TRPA1(B) (Fig. 4b, c). Thus, although TRPA1(B)-specific sequences are unnecessary for robust responsiveness to innocuous warming, they may tune channel threshold within this range.

In insect disease vectors, TRPA1 orthologues have been implicated in detecting both warmth and chemical repellents[4,23,24], cues with opposing effects on host seeking. We found the malaria mosquito *Anopheles gambiae* also contains TRPA1(A) and TRPA1(B) isoforms of differing thermosensitivity (Fig. 4d–h). In oocytes, the Q_{10} of *A. gambiae* TRPA1(A) was ~4 versus a Q_{10} of ~200 for *A. gambiae* TRPA1(B); from 27–37 °C, the Q_{10} of *A. gambiae* TRPA1(A) was ~2 versus ~60 for *A. gambiae* TRPA1(B) (Fig. 4h). *A. gambiae* TRPA1(A) yielded lower maximum heat-induced current than *A. gambiae* TRPA1(B) (Supplementary Fig. 4) and had a higher threshold

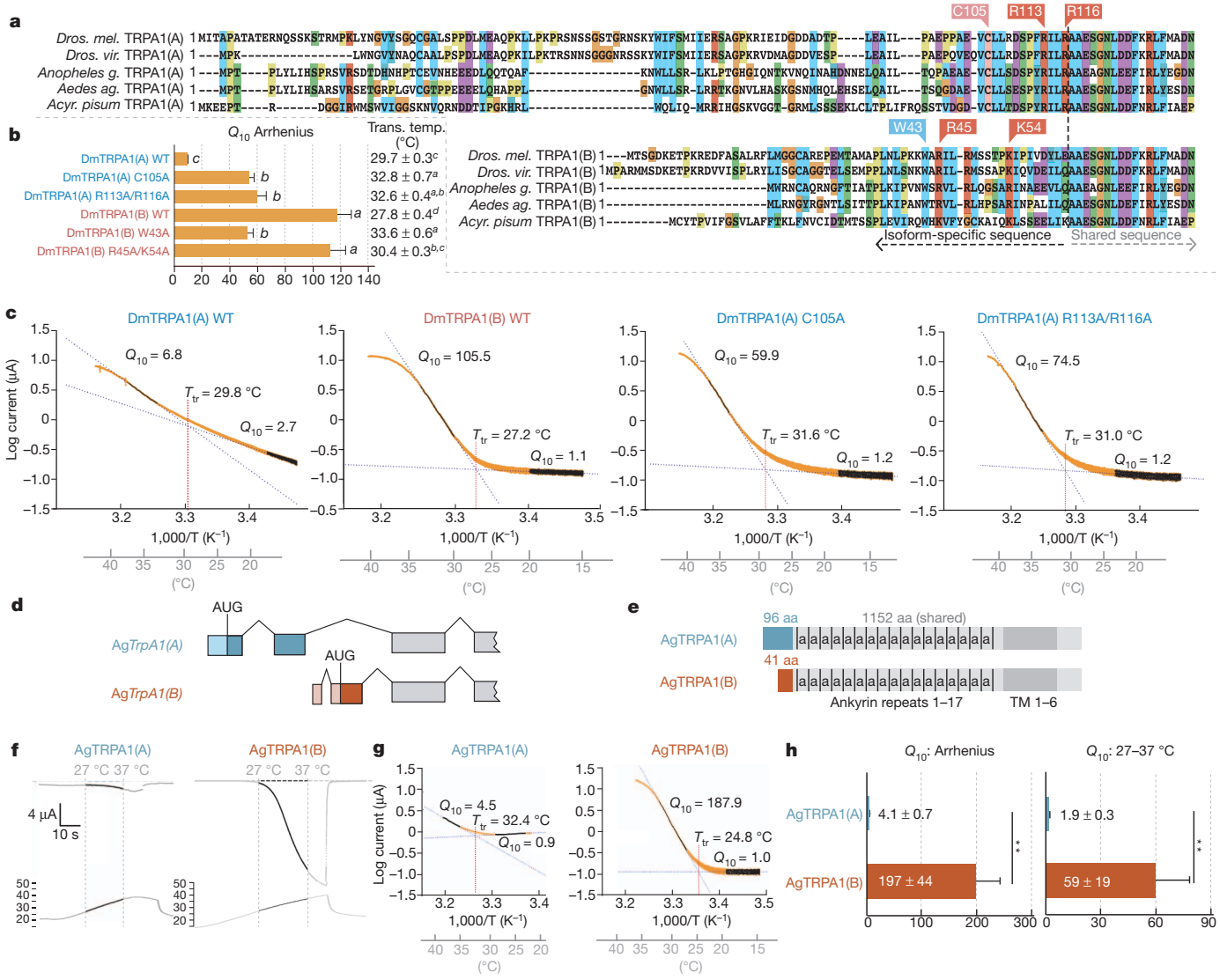


Figure 4 | Regulation of insect TRPA1 thermosensitivity by alternative N termini. **a**, TRPA1 sequence alignments. *Acyr. pisum*, *Acyrthosiphon pisum*; *Aedes ag.*, *A. aegypti*; *Anopheles g.*, *A. gambiae*; *Dros. mel.*, *D. melanogaster*; *Dros. vir.*, *D. virilis*. **b**, Q_{10} values and transition temperatures (Trans. temp.) of wild-type (WT) and mutant TRPA1. Dm, *D. melanogaster*. *a*, *b* and *c* denote statistically distinct groups (Tukey HSD, $\alpha = 0.02$). **c**, Arrhenius plots of indicated channels. **d**, *A. gambiae* (Ag)*TrpA1* gene structure. **e**, AgTRPA1 isoforms. a, ankyrin repeat; aa, amino acids. Blue and dark red indicate isoform-specific amino acids. Dark grey indicates transmembrane (TM) region. **f–h**, Temperature sensitivity of AgTRPA1(A) and AgTRPA1(B). **f**, **g**, Traces (**f**) and Arrhenius plots (**g**) of temperature-dependent current recordings at −60 mV in *Xenopus* oocytes. **h**, Q_{10} values from Arrhenius plot (left) or 27–37 °C (right) (**$P < 0.01$, *t*-test). Error bars, ± s.e.m.

(34.2 ± 1.8 °C versus 25.2 ± 0.9 °C, $P < 0.01$). *A. gambiae* TRPA1(A) also exhibited significant conductance below threshold (Fig. 4f, g). Both channels responded to electrophiles (Supplementary Fig. 9). TRPA1(A) and TRPA1(B) are conserved in other haematophagous insects including *Aedes aegypti* and *Culex quinquefasciatus* mosquitoes and *Pediculus humanus corporis* lice (Fig. 4a and Supplementary Fig. 10), which transmit dengue, West Nile fever and typhus, respectively. The functional diversity of TRPA1 provides a potential explanation for how insect vectors discriminate noxious chemicals from host-derived warmth, indicating that TRPA1 presents two distinct molecular targets for disrupting pest behaviour.

TRPA1-based electrophile detection appears to have emerged ≥500 million years ago in a common vertebrate/invertebrate ancestor[4]. However, the larger TRPA family extends to choanoflagellates, separated from animals ≥600 million years[4]. As divergent TRPA clades contain highly temperature-sensitive channels[1], thermosensitivity may be ancestral. In this scenario, the specialization of TRPA1 for noxious chemical detection would necessitate reducing thermosensitivity, consistent with the effect of the N terminus in TRPA1(A). The ability of N-terminal variation to sculpt channel properties is intriguing as the N terminus is the most divergent region of TRPA1 within insects and from insects to mammals[5].

TRPs are a large family of channels, with 27 human and 13 *Drosophila* members, which vary greatly in thermosensitivity and function[2]. Considerable diversity is evident even among closely related TRPs. In mammals, for example, TRPM8 (Q_{10} ~24; ref. 21) mediates thermosensation[3], whereas the less thermosensitive TRPM4 and TRPM5 (Q_{10} ~8.5–10; ref. 15) mediate insulin secretion[25] and TRPM7 (with no reported thermal sensitivity) is implicated in ion homeostasis[2]. The mechanisms underlying such diversification are unclear. Whereas studies of thermal sensing by TRPs have focused on identifying regions promoting thermosensitivity[1,26–29], our work indicates that regions reducing thermosensitivity are also critical. Here we find that selectively reducing the thermosensitivity of

Drosophila TRPA1 facilitates its use in a context in which thermosensitivity is undesirable. Similar mechanisms could mediate functional diversification not only among isoforms of a single TRP, but also contribute to the remarkable functional diversification observed between different TRP family channels.

METHODS SUMMARY

Fly strains and immunohistochemistry. *UAS-TrpA1(B)* and *Gr66a-Gal4* transgenic strains and the $TrpA1^{ins}$ mutant have been described[4]. The *UAS-TrpA1(A)* transgene was amplified from fly complementary DNA with an isoform-specific primer (5′-TATAAAGCTTAAGCCACCATGATTACAGCTCCGGCCACGGCCA-3′) and a reverse primer (5′-GAGACTCGAGCTACATGCTCTTATTGAAGCTCAGGGCG-3′). As detailed in Methods, the *UAS-TrpA1(A)* transgene was inserted in the same genomic location used for the *UAS-TrpA1(B)* transgene to control for transgene position effects. Anti-TRPA1 immunohistochemistry was as described[4], except that the secondary antibody was incubated for 3 days.

Behaviour. The proboscis extension assay was conducted as previously described[4], with seven flies per experiment, three experiments per genotype. For heat-sensitive regurgitation, >20 flies per genotype (2–3 days old) were starved overnight with water, then glued to glass slides. After 2–3 h recovery, flies were satiated with water. Only flies drinking longer than 5 s were tested. Drinking times did not significantly differ between wild-type and rescue flies (E.C.C. and P.A.G., unpublished data), consistent with similar ingestion behaviours. Flies were heated with a radiant heater at 800 W (H-4438, Optimus) and temperature was monitored by adjacent thermocouple microprobe (IT-23, Physitemp Instruments) wrapped in fly cuticle.

Physiology. Oocyte physiology was performed as described[4], with additional details provided in Methods. Extracellular recordings of gustatory neurons were obtained by tip recording[9], as detailed in Methods.

Molecular biology. cDNA was prepared from dissected *D. melanogaster* tissue (RETROscript, Ambion) and subjected to PCR with three primers, two forward isoform-specific primers and a reverse common primer detailed in Methods. Each primer was designed to straddle a splice junction, minimizing amplification of contaminating genomic DNA. In all cases, similar results were obtained from four independent tissue preparations.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 September; accepted 16 November 2011.
Published online 4 December 2011.

1. Dhaka, A., Viswanath, V. & Patapoutian, A. TRP ion channels and temperature sensation. *Annu. Rev. Neurosci.* **29,** 135–161 (2006).
2. Wu, L. J., Sweet, T. B. & Clapham, D. E. International Union of Basic and Clinical Pharmacology. LXXVI. Current progress in the mammalian TRP ion channel family. *Pharmacol. Rev.* **62,** 381–404 (2010).
3. Daniels, R. L. & McKemy, D. D. Mice left out in the cold: commentary on the phenotype of TRPM8-nulls. *Mol. Pain* **3,** 23 (2007).
4. Kang, K. *et al.* Analysis of *Drosophila* TRPA1 reveals an ancient origin for human chemical nociception. *Nature* **464,** 597–600 (2010).
5. Hamada, F. N. *et al.* An internal thermal sensor controlling temperature preference in *Drosophila*. *Nature* **454,** 217–220 (2008).
6. Patapoutian, A., Tate, S. & Woolf, C. J. Transient receptor potential channels: targeting pain at the source. *Nature Rev. Drug Discov.* **8,** 55–68 (2009).
7. Viswanath, V. *et al.* Opposite thermosensor in fruitfly and mouse. *Nature* **423,** 822–823 (2003).
8. Cohet, Y. Epigenetic influences on the lifespan of the *Drosophila*: existence of an optimal growth temperature for adult longevity. *Exp. Gerontol.* **10,** 181–184 (1975).
9. Hodgson, E. S., Lettvin, J. Y. & Roeder, K. D. Physiology of a primary chemoreceptor unit. *Science* **122,** 417–418 (1955).
10. Weiss, L. A., Dahanukar, A., Kwon, J. Y., Banerjee, D. & Carlson, J. R. The molecular and cellular basis of bitter taste in *Drosophila*. *Neuron* **69,** 258–272 (2011).
11. von Philipsborn, A. C. *et al.* Neuronal control of *Drosophila* courtship song. *Neuron* **69,** 509–522 (2011).
12. Graveley, B. R. *et al.* The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471,** 473–479 (2011).
13. Vyklicky, L. *et al.* Temperature coefficient of membrane currents induced by noxious heat in sensory neurones in the rat. *J. Physiol. (Lond.)* **517,** 181–192 (1999).
14. Tracey, W. D. Jr, Wilson, R. I., Laurent, G. & Benzer, S. *painless*, a *Drosophila* gene essential for nociception. *Cell* **113,** 261–273 (2003).
15. Talavera, K. *et al.* Heat activation of TRPM5 underlies thermal sensitivity of sweet taste. *Nature* **438,** 1022–1025 (2005).
16. Gracheva, E. O. *et al.* Molecular basis of infrared detection by snakes. *Nature* **464,** 1006–1011 (2010).
17. Gracheva, E. O. *et al.* Ganglion-specific splicing of TRPV1 underlies infrared sensation in vampire bats. *Nature* **476,** 88–91 (2011).
18. Marella, S. *et al.* Imaging taste responses in the fly brain reveals a functional map of taste category and behavior. *Neuron* **49,** 285–295 (2006).
19. Lu, G., Henderson, D., Liu, L., Reinhart, P. H. & Simon, S. A. TRPV1b, a functional human vanilloid receptor splice variant. *Mol. Pharmacol.* **67,** 1119–1127 (2005).
20. Vos, M. H. *et al.* TRPV1b overexpression negatively regulates TRPV1 responsiveness to capsaicin, heat and low pH in HEK293 cells. *J. Neurochem.* **99,** 1088–1102 (2006).
21. Brauchi, S., Orio, P. & Latorre, R. Clues to understanding cold sensation: thermodynamics and electrophysiological analysis of the cold receptor TRPM8. *Proc. Natl Acad. Sci. USA* **101,** 15494–15499 (2004).
22. Liu, B., Hui, K. & Qin, F. Thermodynamics of heat activation of single capsaicin ion channels VR1. *Biophys. J.* **85,** 2988–3006 (2003).
23. Wang, G. *et al. Anopheles gambiae* TRPA1 is a heat-activated channel expressed in thermosensitive sensilla of female antennae. *Eur. J. Neurosci.* **30,** 967–974 (2009).
24. Maekawa, E. *et al.* The role of proboscis of the malaria vector mosquito *Anopheles stephensi* in host-seeking behavior. *Parasit. Vectors* **4,** 10 (2011).
25. Uchida, K. & Tominaga, M. The role of thermosensitive TRP (transient receptor potential) channels in insulin secretion. *Endocr. J.* doi:10.1507/endocrj.EJ11-0130 (2011).
26. Grandl, J. *et al.* Pore region of TRPV3 ion channel is specifically required for heat activation. *Nature Neurosci.* **11,** 1007–1013 (2008).
27. Grandl, J. *et al.* Temperature-induced opening of TRPV1 ion channel is stabilized by the pore domain. *Nature Neurosci.* **13,** 708–714 (2010).
28. Yang, F., Cui, Y., Wang, K. & Zheng, J. Thermosensitive TRP channel pore turret is part of the temperature activation pathway. *Proc. Natl Acad. Sci. USA* **107,** 7083–7088 (2011).
29. Yao, J., Liu, B. & Qin, F. Modular thermal sensors in temperature-gated transient receptor potential (TRP) channels. *Proc. Natl Acad. Sci. USA* **108,** 11109–11114 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the P.A.G. laboratory, H. Garrity, L. Griffith, L. Huang, J. Rodriguez and M. Rosbash for helpful comments, and F. Marion-Poll and A. Dahanukar for guidance with tip-recording. This work was supported by grants from the National Science Foundation (IOS-1025307), National Institute of Mental Health (EUREKA R01 MH094721) and National Institute of Neurological Disorders and Stroke (NINDS) (PO1 NS044232) to P.A.G., a National Research Service Award from NINDS to V.C.P. (F31 NS071897-02) and the Boston College DeLuca Professorship to M.A.T.M.

Author Contributions: K.K., V.C.P. and P.A.G. designed experiments. K.K. performed molecular biology, genetics and oocyte physiology. V.C.P. performed genetics and sensory neuron electrophysiology. E.C.C. performed genetics and behavioural experiments. A.M.D. performed behavioural experiments. L.N. performed immunohistochemistry. A.M.J., K.R. and M.A.T.M. grew and harvested mosquitoes. P.A.G. performed bioinformatics. K.K., V.C.P. and P.A.G. wrote the paper.

Author Information TRPA1(A) sequence is deposited in GenBank under accession number JQ015263. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.A.G. (pgarrity@brandeis.edu).

METHODS

Fly strains and immunohistochemistry. The *UAS-TrpA1(A)* transgene was inserted into the genome by site-specific transgenesis[30] at the same landing site as *UAS-TrpA1(B)*, attp16 (ref. 31).

Behavioural analysis. Chemicals used in incapacitation assays were sucrose (Calbiochem LC8510), sorbitol (Sigma S-1876), ficoll (Sigma F-4375), agarose (Invitrogen 15510-027), caffeine (Sigma C0750), NMM (Sigma 389412), isopropanol (100%, J. T. Baker 9083-03), ethanol (100%, Decon Lab 2716) and allyl isothiocyanate (95%, Sigma 377430).

Characterization of TRPA1 isoforms in *Xenopus* oocytes. TRPA1 currents were recorded as described[4,5]. To evaluate temperature sensitivities, oocytes were perfused in the recording buffer (96 mM NaCl, 1 mM $MgCl_2$, 4 mM KCl, and 5 mM HEPES, pH 7.6), the temperature of which was increased $\sim$0.5 °C s^{-1} from 10 to 45 °C by SC-20 in-line heater/cooler (Warner Instruments) with a CL100 bipolar temperature controller (Warner Instruments). Temperature-evoked current was recorded at −60 mV. From the recorded current, Q_{10} was calculated as described[13,16]. Arrhenius $Q_{10} = 10^{(10(-S_{arrhe})/(T_1T_2))}$, where S_{arrhe} is the slope of linear phase of an Arrhenius plot between absolute temperatures, T_1 and T_2. Transition temperature was assessed as the temperature at which the least-squares fit lines from the two linear phases intersect[13,16]. Q_{10} from 27–37 °C was calculated from currents at temperatures of interest using the equation, $Q_{10} = (I_2/I_1)^{10/(T_2-T_1)}$, where I_1and I_2 are currents observed at temperatures of T_1 and T_2, respectively. Q_{10} determinations were validated by using *Crotalus atrox* TRPA1 (ref. 16) as a control with known Q_{10} (K.K. and P.A.G., unpublished data). To assess sensitivity to NMM, voltage across the membrane was initially held at −80 mV, and a 300-ms voltage ramp (−80 mV to 80 mV) per second was applied. The oocytes were perfused for 1 min with the recording buffer containing indicated concentrations of NMM with 30-s washes between NMM applications. Current amplitudes at −80 mV after application of each NMM concentration were fitted to the Hill equation through Sigmaplot 10. The first coding exon of *A. gambiae TrpA1(B)* was chemically synthesized (Genscript).

Gustatory neuron electrophysiology. Extracellular recordings of gustatory neurons were obtained using the tip-recording method[9]. Adult female flies, aged 1–4 days, were prepared by inserting a glass reference electrode containing *Drosophila* Ringer's solution into the thorax and advancing the electrode through the head to the labellum. A glass recording electrode with an $\sim$15-μm opening was used to apply tastants to individual sensilla. Raw signals were amplified using a TasteProbe preamplifier (Syntech) and were digitized and analysed using a PowerLab data-acquisition system with LabChart software (ADInstruments). Amplified signals were digitized at a rate of 20 kb s^{-1} and filtered using a 100–3,000 Hz band-pass filter before analysis. Individual action potentials were sorted using a visually adjusted threshold and average spike rate was calculated beginning 200 ms after electrode contact. Recording times varied by experiment: berberine chloride and sucrose positive controls, 5 s; electrolyte only, 20 s; NMM on i-type bristles, 60 s; NMM on L-type bristles, 120 s; heat ramps, >60 s. For heat-ramp experiments, recordings were performed using electrolyte only as tastant. After $\sim$30 s of recording to determine baseline activity, heat was applied manually to the fly using a radiant heater (PRESTO HeatDish, National Presto Industries). Application of heat was maintained for $\sim$10–30 s and the distance between the heat source and the preparation was reduced to obtain a temperature of ≥39 °C. Bristle temperature was estimated using thermocouple microprobe (IT-23, Physitemp Instruments) wrapped in fly cuticle. All tastants were dissolved in 30 mM tricholine citrate as the electrolyte to inhibit the activity of the water cell in L-type bristles[32]. Tastants were stored at −20 °C and aliquots maintained at 4 °C for up to 1 week. For all experiments, a positive control was used to confirm the viability of the target bristle. For i-type bristles, 1 mM berberine chloride was used as control. For L-type bristles, 30 mM sucrose was used. Individual tastant presentations were separated by a minimum delay of 60 s. At least two animals and six bristles were examined for each condition.

Molecular biology. Primers for RT–PCR reactions: *D. melanogaster TrpA1(A)* forward, 5′-GCCGGAACAGCAAGTATT3-3′; *D. melanogaster TrpA1(B)* forward, 5′-GTGGACTATCTGGAGGCG-3′; *D. melanogaster TrpA1* common reverse, 5′-TATCCT TCGCATTAAAGTCGC-3′.

Mutagenesis of *Drosophila* TRPA1 was performed as described[4]. Briefly, for a desired mutation, each of two mutually complementary mutant primers was paired for PCR with a primer (outer primer) that anneals outside of either SalI or HpaI restriction recognition site. The two resulting PCR fragments that overlap only in the region of the two mutant primers were combined and served as template for the next PCR reaction that contained only outer primers. The second PCR product was digested by SalI and HpaI, and subsequently replaced the corresponding wild type region of TRPA1 cDNA. The fragment between the two restriction sites was sequenced. Sequences were aligned using MUSCLE 3.7[33].

30. Groth, A. C., Fish, M., Nusse, R. & Calos, M. P. Construction of transgenic *Drosophila* by using the site-specific integrase from phage ϕC31. *Genetics* **166,** 1775–1782 (2004).
31. Markstein, M., Pitsouli, C., Villalta, C., Celniker, S. E. & Perrimon, N. Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nature Genet.* **40,** 476–483 (2008).
32. Wieczorek, H. & Wolff, G. The labellar sugar receptor of *Drosophila*. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **164,** 825–834 (1989).
33. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).